Università degli Studi di Pisa
Scuola Superiore Sant'Anna di Studi Universitari e Perfezionamento

DIPARTIMENTO DI ECONOMIA E MANAGEMENT
Corso di Laurea Magistrale in Economics

Tesi di laurea magistrale

Short-Run Effects of the Abolition of Health User Fees: Evidence from the Free Health Care Initiative in Sierra Leone

Laureanda:
Eleonora Guarnieri

Relatori:
Prof. Simone D'Alessandro
Prof.ssa Katja Maria Kaufmann

Anno Accademico 2014–2015

Abstract

I exploit a policy change in Sierra Leone to investigate the effect of the elimination of health user fees on the demand for women and children use of medical care and on health outcomes, such as anthropometric indicators and mortality. I use the Demographic and Health Survey dataset including the aforementioned information for mothers and children. The reform I study came into effect in April 2010 and it abolished hospital user fees for pregnant and lactating women and children under five. For each indicator of interest I estimate a difference-in-differences model using older cohorts of children or women delivering before the policy change as controls. I find significant effects on the frequency of women delivering in a health facility and on weight-for-age for children under five. I do not identify any effect for child mortality, whereas effects on maternal mortality are ambiguous. Effects vary substantially by income, place of residence and distance to the health facility, with women and children with middle and high income, urban and far away households being in general more responsive to the policy change. In order to add credibility to the results I relax the functional form assumption and repeat the estimation non-parametrically through a kernel weighted propensity score matching strategy. My results shed some light on the potentially positive short-run effects of reducing financial barriers to the access of health care in low income countries.

Acknowledgements

Firstly, I would like to express my gratitude to my supervisors. To Prof. Katja Maria Kaufmann at the University of Mannheim, for her constant guidance, patience and deep knowledge. To Prof. Simone D'Alessandro at the University of Pisa, for his support throughout the last phase of the realization of this thesis and for his always encouraging attitude.
My sincere thanks also goes to Prof. Juanita Vasquez-Escallon, for her insightful advice.
I want to thank Elena, whose wisdom guided me during the two years of my Master.
Thanks to Sebastian, Kemal, Bianca and Matze, who filled with joy my year in Mannheim.
Thanks to Veronica, for accompanying me through the last ten years of my life with patience and empathy.
Thanks to Raffaele, and to all my friends.
Finally, all my gratitude goes to my mother, my father and Giovanni, for their endless love.

Contents

List of Figures

List of Tables

List of Abbreviations

AIDS Acquired Immuno-Deficiency Syndrome
CHC Community Health Centers
CHP Community Health Posts
DHS Demographic and Health Survey
EA Enumeration Area
FHCI Free Health Care Initiative
GIS Geographic Information System
GPS Global Positioning System
HIV Human Immunodeficiency Virus
ITN Insecticide Treated Net
MCHP Maternal and Child Health Post
MDGs Millennium Development Goals
MMR Maternal Mortality Ratio
NCHS National Center of Health and Statistics
WHO World Health Organization
PHU Peripherial Health Unit
PSM Propensity Score Matching
PSU Primary Sampling Unit
RCT Randomized Controlled Trial
SECHN State Enrolled Community Health Nurse
SUTVA Stable Unit Treatment Value Assumption
TBA Traditional Birth Attendants

Introduction

Starting from the 1980s and during the 1990s, the priority for policy makers in developing countries and consequently the attention of the literature was mainly focused on finding measures for improving the quality of health care facilities, worsened by a crisis in health care financing. This need was a consequence of a considerable increase in demand for services, due to general population growth and to the appearance of new health problems, such as the HIV and AIDS (McPake 1993). One of the packages of measures put forward by policymakers, encouraged by international institutions such as the World Bank, was the introduction of user fees as an immediate solution to the problem and as preparation for the future implementation of other options such as health insurance schemes. Among the advantages of this approach, one can identify the room for quality improvements ensured by a considerable increase in revenues and thus the possibility of welfare gains for the majority. On the other hand, potential problems such as lack of equity in the opportunity to benefit from health service and a decrease in utilization have been widely discussed.

With the new millennium, priorities for the developing world, as far as health is concerned, have moved to other directions. The United Nations Millennium Development Goals state the importance of reducing child mortality and improving maternal health, together with fighting HIV/AIDS, malaria and other diseases. A channel through which these aims are seen to be achieved is encouraging the access to health care through the reduction or elimination of health care user fees for targeted sensitive groups, like mothers and children. In poor developing countries like Sierra Leone, direct costs such as fees constitute an important barrier to the access even to basic health care services. Moreover, traditional (and unskilled) healers continue to be perceived as an affordable and trustworthy source of care, especially among groups in poverty conditions and unable to reach any health facility because living in remote rural areas. Access to health care is

thus seen as the first step to be implemented for achieving the aforementioned broader goals.

The aim of this study is providing some empirical evidence to assess the validity of a measure like removing user fees on goals such as increasing incentives to seek necessary care and improving health conditions for sensitive categories like mothers and children. The context under investigation is Sierra Leone, one of the poorest countries in the world, and its recent introduction of the Free Health Care Initiative, which eliminated user fees for pregnant and lactating women and children under five and endowed hospitals with additional drugs and machinery. The policy change I study came into effect in April 2010. Before the reform, children and mothers were subject to user fees that were ranging between 5,000 Le (1.15$) for consultation and 300,000 Le (69.04$) for a complex major surgery.

I use the Sierra Leone Demographic and Health Survey dataset for 2013, which includes some variables of interest such as date of birth of children, anthropometric indicators, place of delivery and mortality, together with socioeconomic variables such as wealth, education and place of residence. The large sample size and the possibility to link the outcome variables to a specific month and year make this dataset ideal for the purpose of the study.

My analysis employs a difference-in-differences design, where the control group includes alternatively cohorts of children born in 2009, when the policy was not into effect, or mothers that delivered between 2004 and 2009, depending on the outcome variable of interest. The treatment group includes children born in 2010 and mothers delivering in 2010. The identifying assumption is that treatment and control groups exhibit similar trends in the outcome variable over time, which in this setting translates into the assumption that there would be similar seasonality effects for each group, in absence of the policy (this is the reason why groups are selected in a way that they cover whole years). One way I adopt to indirectly test this hypothesis, which is of course not testable in itself, is checking whether there is balance between the groups in terms of some observable characteristics. Moreover, I investigate whether there is some heterogeneity in the responsiveness to the program for groups in rural or urban areas, belonging to lower or higher wealth quintiles, living far or close to a hospital, being more or less educated and being male or female, in the case of anthropometric indicators.

My results show that abolishing health user fees for children and mothers increases

the probability that a woman delivers in a governmental health facility by roughly 4 percentage points. This result is similar to the one predicted by a nonparametric specification. The effect is higher for richer women and for women living in urban areas, whereas there seem to be a higher effect for groups of women living further away from the health facility. The policy might have encouraged to find ways to reach a hospital for delivering and to be willing to face some indirect cost, given the absence of direct costs.

The analysis shows short run positive effects on health outcomes for children, captured by anthropometric indicators. The ones included in my analysis are weight for height, height for age and weight for age. Significant effects are seen only for the latter, with higher effects for males, and, at odds with the place of delivery, for poor children. Ambiguous short-term beneficial effects are recorded for maternal mortality. It seems to have decreased of roughly 2,000 per 100,000 live births, but the results are not statistically significant. On the other hand, the analysis does not show a clear pattern for child mortality.

The contribution of this study is to show how reducing user fees in the context of Sierra Leone increases demand for health service, as revealed by the choice of mothers to deliver in a governmental health center, instead of relying on traditional birth attendants. This confirms the tendency of other countries in which a similar policy was implemented, such as Ghana and Uganda (Penfold et al. 2007, Asante et al. 2007, Arthur 2012). Moreover, the policy is shown to have had positive effects in terms of general health conditions for children and a desirable effect, even though not significant, for maternal mortality. Instead, as pointed out in other similar studies (Bosu et al. 2007) no short run desirable effects on child mortality are recorded. This might be due to a mother biased attitude in hospitals or due to the fact that this outcome is responsive to the policy only in a long run horizon.

The different degree of responsiveness among poor children suggests to policymakers that the removal of health user fees can contribute to encourage access to health care to sensitive categories, for which the price elasticity of demand is shown by some studies to be particularly high (Gertler et al. 1987, Ching 1995), thus reducing inequality. However, the fact that the effect on the place of delivery is more pronounced for richer mothers leaves open the question about whether removing economic barriers to health care is enough to encourage access.

The rest of the thesis is organized as follows. Chapter 1 outlines some of the main

empirical methods adopted when evaluating policies or programs. Chapter 2 provides an overview of previous studies on the introduction or removal of user fees in many African and non-African developing countries. Chapter 3 proposes the evaluation of the Free Health Care Initiative in Sierra Leone.

Chapter 1

Empirical Methods for Impact Evaluation

In the process of recharging growth and reducing poverty, increasing attention by governments, aid donors and international institutions has been devoted to generate empirical evidence in order to assess which policies are effective in reaching such goals. To this end, researchers have sought scientific tools and developed evaluation approaches in order to implement precise estimation of the effects, conferring not only credibility to the results, but also providing useful insights for policy makers in developing countries. This has allowed to answer some crucial questions, such as which are the most effective and efficient policies in different contexts or which are the best ways to implement specific programs.

The aim of this chapter is to provide an overview of the most widespread empirical methods adopted in evaluating the effect of policies like the ones discussed in the following chapters, namely the introduction or removal of health user fees. These kind of policies fall into the category of *natural* experiments, as defined in Rosenzweig & Wolpin (2000) and Meyer (1995). After an overview of the archetypal formulation of the so-called evaluation problem, the following sections will clarify the issues involved in identifying the impact of such policies, the problems that a researcher is likely to face and some techniques for overcoming such difficulties[1].

[1]The main sources for this chapter are Ravallion (2007), Duflo (2002) and Khandker et al. (2010).

1.1 The Identification Problem

The ultimate goal of an impact evaluation analysis is to assess a program or policy performance in attaining some specified objectives against an explicit counterfactual, such as the absence of the program. In other words, empirical methods have been developed in order to answer counterfactual or *what if* questions, such as: *what would have happened to individual i if he or she had not benefited from the policy or if he or she had been exposed to an alternative one?*

This problem can be formalized as follows. Given a specific outcome of interest Y, what researchers are interested in is the change in such outcome that can be causally attributed to a particular program or policy, the so-called *impact*. The researcher observes such outcome for a sample of individuals. Ideally, for each individual i in the sample, two are the outcomes of interest: Y_i^T, the outcome under treatment, Y_i^C, the outcome of the same individual in the absence of the program, the so-called counterfactual outcome. The impact (or causal effect), which is what one wants to identify, can thus be defined as:

$$Y_i^T - Y_i^C \tag{1.1}$$

Unfortunately, the same individual cannot be observed as being exposed to the program and as not being exposed at the same time. This is what the literature has referred to as the *missing data problem* (Ravallion 2007): it is physically impossible to measure outcomes for an individual simultaneously in two different states of nature. Moreover, what one can aim at identifying is not an effect on a single individual i, but instead the expected gain over a population as follows:

$$E[Y_i^T - Y_i^C] \tag{1.2}$$

One possible and straightforward way of overcoming the missing data problem is taking into consideration a group of individuals that are not exposed to the program, whose outcome is thus likely to approximate Y_i^C. This group of individuals is often referred to as the *control group*, in order to distinguish it from the *treatment group*, the set of individuals exposed to the policy or program. Given these definitions, the naïve estimator of the impact or treatment effect (D) can be expressed as follows:

$$D = E[Y_i^T|T] - E[Y_i^C|C] \tag{1.3}$$

Both elements in Equation 1.3 are observed, therefore the treatment effect can be estimated. By adding and subtracting $E[Y_i^C|T]$, i.e. the not-observed counterfactual outcome, one obtains:

$$\begin{aligned} D =& E[Y_i^T|T] - E[Y_i^C|T] - E[Y_i^C|C] + E[Y_i^C|T] = \\ =& \underbrace{E[Y_i^T - Y_i^C|T]}_{\text{Impact}} + \underbrace{E[Y_i^C|T] - E[Y_i^C|C]}_{\text{Selection Bias}} \end{aligned} \tag{1.4}$$

The element of interest is the impact, summarized in the first term of Equation 1.4 (specifically called *treatment effect on the treated*). However, what can be noticed is that the naïve estimator of the treatment effect (D) can coincide with the actual impact ($E[Y_i^T - Y_i^C|T]$) only when the second term of Equation 1.4 is zero. The difference $E[Y_i^C|T] - E[Y_i^C|C]$ is known as *selection bias*, which can be described as an underlying difference in the outcome Y not reconcilable to the exposition to the program. The selection bias can arise because individuals could be assigned to treatment or control on the basis of some observed or unobserved characteristics, that makes the difference in outcome be systematic and not necessarily linked to the policy under investigation.

Another possible design for the identification of the treatment effect is a before-and-after estimator that compares the outcome for the treatment group before and after the implementation of the policy:

$$D = E[Y_1|T] - E[Y_0|T] \tag{1.5}$$

where Y_0 refers to the period before the policy and Y_1 to the period after the policy. In order for this to be an unbiased estimator, there are two strict assumptions that have to be fulfilled: there should not be time-variant effects that have an impact on the outcome Y other than the policy itself; the pre-treatment outcome should not be affected by the policy or program (for example, in the case in which some requirements are needed in order to have access to the program). This is clear by observing Figure 1.1. Assuming that the outcome of interest is income (on the Y axis), the just mentioned approach would compute the treatment effect as $Y_2 - Y_0$. However, the figure shows through the dashed line that if income is affected by a trend which is not directly related to the program,

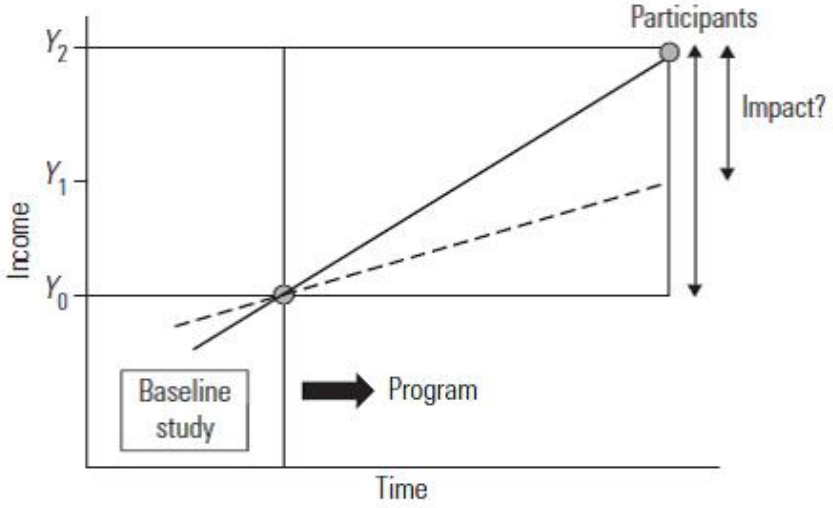


Figure 1.1: Before-and-After Comparison. Source: Khandker et al. (2010)

then the treatment effect thus computed would overestimate the actual impact $(Y_2 - Y_1)$.

Several empirical methods have been developed in order to solve the identification problem. Each of these methods carries its own assumptions about the nature of potential selection bias in program targeting and participation, and the assumptions are crucial to developing the appropriate model to determine program impacts. Among the various existing methods, only the ones strictly related to the study in Chapter 3 will here be covered: randomized experiments (the ideal research environment), difference-in-differences methods, matching methods (among the latter, specifically propensity score methods).

1.2 Randomized Evaluations versus Natural Experiments

This section describes the ideal research environment to evaluate the effect of a policy X on outcome Y, known as randomized experiment or randomized control trial (RCT). The reason why this is an ideal set-up is that thanks to randomization the selection bias can be assumed to be zero, and the naïve estimator is thus unbiased, for the reasons showed in Section 1.1.

In a randomized experiment, a sample of N individuals is selected from the population, in a way that can also be non-random (for example selected according to observables). This sample is then randomly divided into two groups[2]: the treatment group (composed by N_T individuals) and the control group (composed by N_C individuals). Straightforwardly, $N_T + N_C = N$. N_T individuals are assigned to policy X, whose impact is measured by comparing the outcome Y for treatment and control group as follows:

$$\hat{\mathrm{D}} = \hat{\mathrm{E}}[Y|D] - \hat{\mathrm{E}}[Y|C] \tag{1.6}$$

where $\hat{\mathrm{E}}$ denotes the empirical mean. The reason why randomized experiments represent the ideal research environment, which allows to adopt the naïve estimator, is that the selection bias is by assumption zero, being treatment randomly assigned.

One can easily refer to a regression framework for the estimation of the treatment effect, in order to obtain standard errors for $\hat{\mathrm{D}}$:

$$Y_i = \alpha + D \cdot 1(i \in T) + \epsilon_i \tag{1.7}$$

where $1(i \in T)$ is a dummy variable for belonging to the treatment group and D the parameter of interest.

Despite the possibility to estimate the treatment effect straightforwardly thanks to the absence of selection bias, there are some crucial assumptions that have to be satisfied. Among these, one is the so-called stable unit treatment value assumption (SUTVA): there should not be any interference between the two groups, in the sense that the observed outcome Y_i depends only on treatment, and not on the allocation of other individuals. In other words, there should not be spillover effects. Moreover, one of the possible limitations of RCTs is related to the so-called Hawthrone or John Henry effects: individuals in both groups might behave differently because aware of being observed. These issues make the considerations related to a specific experiment difficult to be extrapolated to other contexts other than the one under study: the experiment is usually limited in time and specific of a country or geographic area. Moreover, the small-scale character of these experiments makes them not likely to generate general equilibrium effects, which might instead be a very important factor in the case the policy was applied on entire popula-

[2]There exist complex sampling designs that involve procedures of stratification, which are not reported here.

tions.

Other shortcomings of randomized evaluations is that experiments tend to be very costly and hard to implement properly. Moreover, some ethical concerns are often expressed by governments when advised to carry out some randomized experiment: the mere distinction between treatment and control group generates inequality in the distribution of benefits.

The just underlined aspects related to randomized experiments can be seen as the reasons why researchers are often not in the condition of designing and implementing an experiment, having instead to evaluate policies developed at a country level, as will be the case in Chapter 3. These are often referred to as *natural experiments*, because treatment and control group are not randomly assigned. For this reason, what researchers have to devote attention to is finding suitable comparison groups in order to answer the aforementioned counterfactual question (Meyer 1995).

Good natural experiments are studies where there is a clear exogenous source of variation in the explanatory variables that determine the treatment assignment, generated by policy changes or other events. The goal for the researcher is to set up an identification strategy for estimating the treatment effect, taking into account the possibility of a selection bias and trying to overcome it by designing a study environment in which treatment and control group are as if they were randomized. The following two sections will be devoted to some techniques that allow the researcher to overcome the identification problem, especially when randomization is not possible.

1.3 Difference-in-Differences Methods

For the reasons highlighted in Section 1.1, both before-and-after and treatment-control designs for the estimation of the treatment effect might turn out to be biased. One widely adopted technique for overcoming both sources of bias is called difference-in-differences strategy, which can be seen as a combination of the two simple difference designs. The basic idea is to compare outcomes for the treatment group to outcomes for the control group both before and after the policy change. This allows to control for both observable and non-observable heterogeneity between the two groups and to keep track of underlying trends in outcome not related to the program. The difference-in-differences estimator is unbiased only under the assumption that the heterogeneity between the two groups is

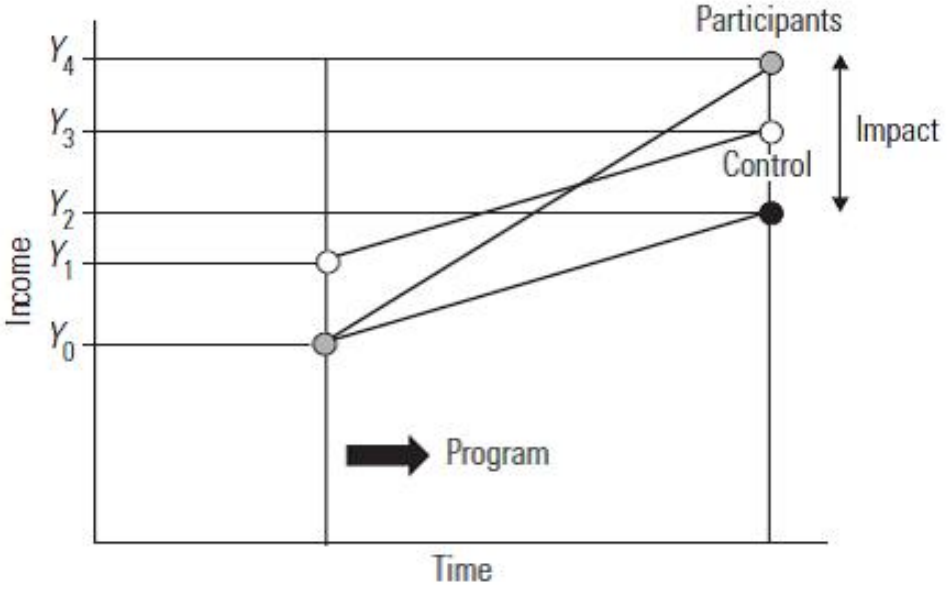


Figure 1.2: Example of Difference-in-Differences. Source: Khandker et al. (2010)

time invariant. In simple words, this means that differences in levels of the outcome between treatment and control group are allowed, whereas differences in trend are not. This is known as the *common trend assumption*: the outcome for the untreated group follows the same trend that, absent the policy change, would have been followed by the group of treated individuals. Analytically, the treatment effect can be expressed as:

$$D_{Did} = [\hat{\mathrm{E}}(Y_1|T) - \hat{\mathrm{E}}(Y_0|T)] - [\hat{\mathrm{E}}(Y_1|C) - \hat{\mathrm{E}}(Y_0|C)] \tag{1.8}$$

Where $\hat{\mathrm{E}}$, again, denotes the empirical mean. The simple difference before-and-after for the treatment group is corrected by subtracting the simple difference for the control group. Equivalently:

$$D_{Did} = [\hat{\mathrm{E}}(Y_1|T) - \hat{\mathrm{E}}(Y_1|C)] - [\hat{\mathrm{E}}(Y_0|T) - \hat{\mathrm{E}}(Y_0|C)] \tag{1.9}$$

Here, the simple difference treatment-control after the policy is corrected by subtracting the simple difference before the policy change. For a visual understanding of the mechanism, one can observe Figure 1.2. The lowermost line represents the true counterfactual, which is in reality never observed. As can be seen, there is a difference in income

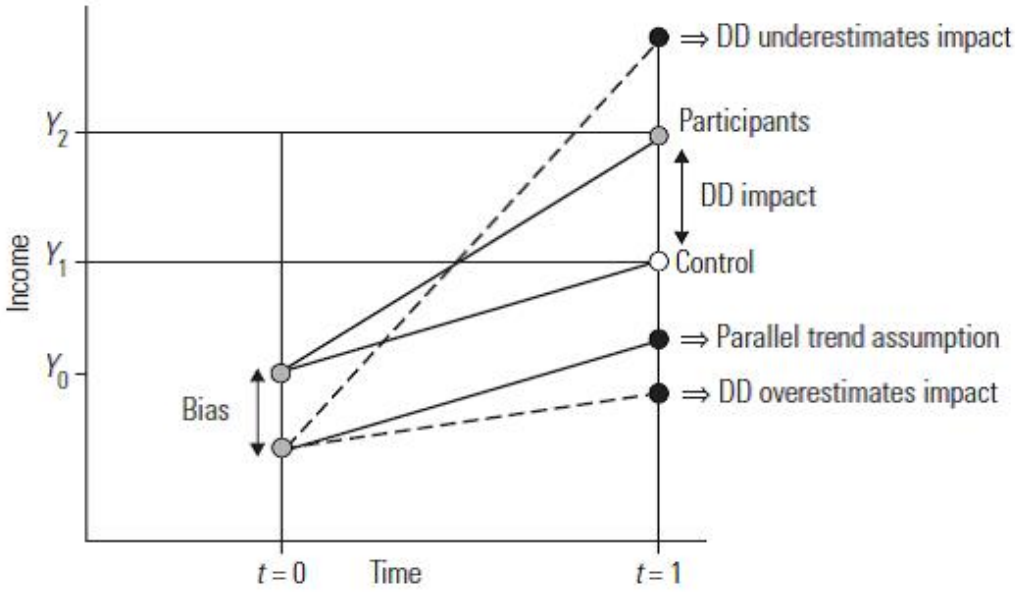


Figure 1.3: Time-Varying Unobserved Heterogeneity. Source: Khandker et al. (2010)

(the outcome of interest) for treatment and control group before the implementation of the program, namely $Y_1 - Y_0$, due to heterogeneity in terms of observable or not-observable characteristics that create the selection bias. Moreover, despite the control group is not affected by the policy, an increasing trend in income can be observed. Assuming that the difference in the baseline is time invariant, the difference-in-differences estimator ($Y_4 - Y_2$, in the case of Figure 1.2) is unbiased in revealing the treatment effect. The regression counterpart of the estimator is the following:

$$Y_{it} = \alpha + \beta \cdot 1(t = 1) + \gamma \cdot 1(i \in T) + \eta \cdot 1(i \in T)1(t = 1) + \epsilon_{it} \tag{1.10}$$

where γ is the parameter of interest, $1(t = 1)$ is a dummy for the period after the implementation of the policy and $1(i \in T)$ is a dummy for treatment.

A shortcoming of this approach occurs whenever the identifying assumption is not satisfied, i.e. in the presence of some time varying unobserved heterogeneity. Figure 1.3 illustrates the point, showing a potential bias when the parallel trend assumption is not satisfied, i.e. when the difference between nonparticipant and (unobserved) counterfactual outcomes changes over time, as illustrated by the dashed lines. Time-varying, unobserved heterogeneity could lead to an upward or downward bias.

1.4 Matching Methods and Selection on Observables

This section illustrates a possible way to overcome the selection bias due to heterogeneity in observable characteristics. Even in the absence of randomization, one can think of constructing a comparison group that looks similar to the group of participants in terms of characteristics that can be observed. This technique is known as *matching*. The core idea of any matching approach is to compare individuals in the treatment group to individuals in the control group which are observationally similar. An immediate shortcoming to this is related to a difficulty in finding similar individuals in terms of an entire vector X of observable characteristics. This is known as the *curse of dimensionality* problem. For this reason, common practice is to adopt the so-called propensity score matching (PSM) approach.

In a nutshell, the practice is to construct a statistical comparison group based on a model of the probability of participating in the program, in turn determined by observed characteristics. Participants are then matched to nonparticipants on the basis of this probability, the *propensity score*, which has the power to incorporate in one measure all the elements of the vector X of observed characteristics. The average treatment effect of the program is then calculated as the mean difference in outcomes across these two groups.

The propensity score is often estimated through a discrete choice model, usually probit or logit. It can be defined as the probability of being treated given some observed characteristics:

$$PS(X) = Pr(T = 1|X) \tag{1.11}$$

where X is a vector of observed characteristics.

Two are the conditions that have to be satisfied in order for this method to be valid. The first is the *conditional independence assumption*, which states that unobserved factors do not affect participation. In other words, potential outcomes for treatment and control

group are independent of treatment once controlling for X[3]:

$$(Y_i^T, Y_i^C) \perp T_i | X_i \tag{1.13}$$

The second assumption is the *common support* or *overlap condition* in propensity score densities across the participant and nonparticipant samples:

$$0 < Pr(T_i = 1|X_i) < 1 \tag{1.14}$$

While the conditional independence assumption is not testable, the latter can be instead tested and this is one of the ways through which one can directly assess whether some selection bias in terms of observable characteristics is likely to be present[4]. The way of assessing the fulfillment of the common support assumption is to compute the propensity score and plot its density for treatment and control group. This is illustrated in Figure 1.4.

In a simple-difference setting, the treatment effect computed using a propensity score matching approach can be expressed as:

$$D_{PSM} = \frac{1}{N_T}[\sum_{i \in T} Y_i^T - \sum_{j \in C} \omega(i,j) Y_j^C] \tag{1.15}$$

where N_T is the number of treated individuals i and $\omega(i,j)$ is the weight used to aggregate outcomes for the matched nonparticipants j.

There are different criteria that can be used to assign participants to nonparticipants on the basis of the propensity score (in other words, different are the forms that the weight $\omega(i,j)$ in Equation 1.15 can assume), and they are usually known as *matching algorithms*[5]. The most common is the kernel matching, the one adopted in Chapter 3. It falls in the category of non-parametric matching estimators, which use a weighted average of all

[3]Rosenbaum & Rubin (1983) show that if potential outcomes are independent of treatment conditional on covariates X, then they are also independent of treatment conditional on a balancing score such as the propensity score. The assumption in Equation 1.13 can be rewritten as:

$$(Y_i^T, Y_i^C) \perp T_i | PS(X) \tag{1.12}$$

where PS(X) is the propensity score as defined in Equation 1.11.

[4]I will adopt this technique in Chapter 3.

[5]Nearest-neighbor, caliper or radius, stratification matching are not covered here.

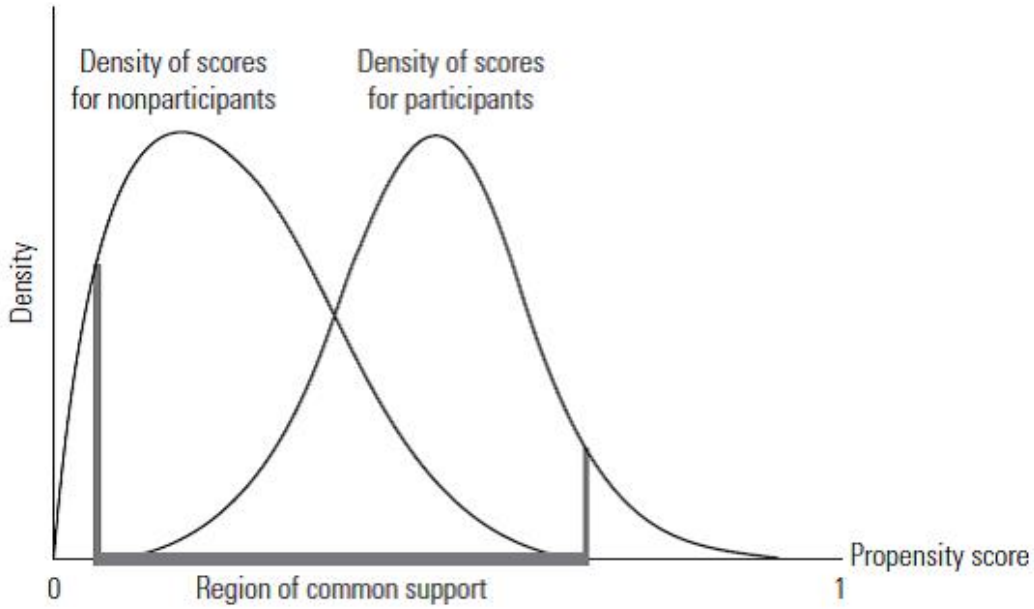


Figure 1.4: Example of Common Support. Source: Khandker et al. (2010)

nonparticipants to construct the counterfactual match for each participant. Defining as P_i the propensity score for participant i and P_j the propensity score for nonparticipant j, the weights for kernel matching are given by:

$$\omega(i,j)_{KM} = \frac{K(\frac{P_j - P_i}{a_n})}{\sum_{k \in C} K(\frac{P_k - P_i}{a_n})} \tag{1.16}$$

where $K(\cdot)$ is a kernel function and a_n is a bandwidth parameter.

One can combine the difference-in-difference approach with kernel-weighted propensity score matching. This allows to account for observed as well as unobserved characteristics affecting participation, always under the assumption that unobserved factors are constant over time. With cross-sections over time rather than panel data[6], the treatment

[6]Only this case is reported, because it reflects the case analyzed in Chapter 3.

effect can be expressed as:

$$D_{PSM}^{DiD} = \frac{1}{N_{T_1}}\left[\sum_{i\in T_1} Y_{i1}^{T} - \sum_{j\in C_1}\omega(i,j)Y_{j_1}^{C}\right] - \frac{1}{N_{T_0}}\left[\sum_{i\in T_0} Y_{i0}^{T} - \sum_{j\in C_0}\omega(i,j)Y_{j_0}^{C}\right] \tag{1.17}$$

where Y_{it}^{T} and Y_{it}^{C} with $t = 1, 2$ are the outcomes for different participants and non-participants observations in each time period t. $\omega(i,j)$, again, represents the weighting function or matching algorithm. In the case of kernel-weighted propensity score matching it coincides with Equation 1.16.

A final note has to be made with respect to inference. Compared to traditional regression methods, the estimated variance of the treatment effect in PSM should include the variance attributable to the derivation of the propensity score and the determination of the common support. Failing to account for this additional variation beyond the normal sampling variation will cause the standard errors to be estimated incorrectly. One solution is to use bootstrapping, where repeated samples are drawn from the original sample and properties of the estimates are re-estimated with each sample. As will be seen, this will be applied in Chapter 3.

The purpose of this chapter was providing an overview on some of the empirical methods commonly used for impact evaluation purposes. It does not aim at being exhaustive, since only the methods adopted in the study in Chapter 3 are mentioned. The following chapter will enter into the specificity of policy interventions concerning health user fees, proposing a theoretical model for an overall understanding and a review of the literature on the topic.

Chapter 2

Health User Fees: Theoretical Framework and Previous Research

The aim of this chapter is highlighting the multiple factors that constitute the core of the widespread debate concerning the role of user fees in developing countries health care systems. This research question gained a first wave of interest in 1980s, when the World Bank encouraged policy-makers in developing countries to consider the implementation of cost-sharing aiming at an affective, efficient, and equitable health care system (World Bank, 1987). A theoretical framework largely exploited by researchers will be presented, in order to encompass the most relevant variables of interest and understand their interaction. Additionally, results of empirical studies based on simulations of this theoretical background will be reported.

A general tendency of the most recent research efforts is basing the empirical strategies on arguments for identifying the causal effect, rather than on complex theoretical backgrounds and on restrictive assumptions. However, the widespread approach of 1980s deserves to be illustrated, and, as said, allows to summarize the conceptual complexity of the issue.

2.1 Behavioural Model of a Household Utility Function in a Health Services Market

An exhaustive theoretical framework for the determinants of health and nutrition and their possible impacts on several outcome variables was proposed by Behrman & Deolalikar (1988). Among the different empirical questions addressed by the model, the one of interest in this context is the following: how responsive is demand for health and consequently health status to changes in prices that households face and in resources that households have? In other words, which is the impact of user fees[1] on access to health care and on health outcomes[2]?

Some of the authors interested in addressing this issue made use of a narrower specification of the aforementioned model, which will be hereby presented, in the version proposed by Gertler et al. (1987)[3].

In short, the model comprises the relationship between household welfare, its socioeconomic characteristics and the attributes of health facilities, such as the price for medical services and their quality. The inclusion of the latter element has to be particularly appreciated. Indeed, as will be stressed in Section 2.3.2, the idea of fees-plus-quality is one of the main arguments in favor of cost-sharing (World Bank, 1987; Litvack and Bodart, 1993). Households are treated as single units, thus excluding from the analysis any intra-household consideration[4]. In this static model, the level of welfare is expressed by a utility function, which in turn depends on health and consumption of goods other than medical care. When any kind of illness arises, the agent faces the problem of choosing whether to seek medical care or not. This economic problem embodies a trade-off: one possible beneficial effect of seeking care is an improvement of expected health status; instead, the cost is captured by a reduction in the consumption of other goods. The choice is not limited to whether to seek care or not, but includes the decision of the kind of health care provider. In order to incorporate these two steps simultaneously, one

[1] User fees are defined as *charges for health care at the point of use* by James et al. (2005).

[2] This was the question of interest in 1980s, when user fees where introduced. This has slowly shifted to the related, even though reverse research question that Chapter 3 will address examining the case of Sierra Leone.

[3] Similar versions can be observed in Mwabu & Mwangi (1986), Mwabu et al. (1993), Ching (1995), Akin et al. (1986), Akin et al. (1995).

[4] Variations in responses within a households are accounted for in Behrman & Deolalikar (1988).

alternative in the choice set will be self-treatment[5]. A price is set by each provider for supplying an improvement in the individual's health status. In order to incorporate both direct and indirect costs of care, the price is comprehensive of direct monetary costs such as user fees and indirect opportunity-costs such as transportation and time, consistently with the idea that these resources would not be available for consumption of other goods. The providers differ one another for their quality, which in this setting translates into different levels of improvement in health. Consistently with economic theory, the rational individual chooses the alternative that yields the highest utility, given the elements above.

In order to confer to the problem a certain degree of formality, let the expected utility of individual i conditional on receiving care from provider j be given by:

$$U_{ij} = U(H_{ij}, C_{ij}, T_{ij}) \tag{2.1}$$

where H_{ij} is the expected health status of individual i after receiving treatment from provider j, C_{ij} is the individual's consumption net of the monetary cost for health care and T_{ij} is the indirect or opportunity cost of access to provider j (such as travel and waiting time). Two further elements conceived to be determinants for the choice of the individual have to be included: quality of health care and income. Quality can be defined as the improvement of the expected health status after the access to provider j with respect to the $j = 0$ choice, which is defined to be self-treatment. Formally, this can be expressed as[6]:

$$Q_{ij} = H_{ij} - H_{i0} \tag{2.2}$$

Alternatively, Equation 2.2 can be rewritten in a way that describes an expected health care production function:

$$H_{ij} = Q_{ij} + H_{i0} \tag{2.3}$$

Quality, as can be noticed by the subscript, is both facility variant and individual

[5] Gertler et al. (1987) include in the analysis a two step choice.

[6] Gertler et al. (1987) provides an alternative formulation, though analogous in the meaning: $Q_{ij} = \frac{H_{ij}}{H_{i0}}$.

variant. If the first feature is straightforward, given that different categories of facilities vary for size, availability of drugs, machinery endowment and staff skills, the relation to the individual deserves particular attention and allows to include an important determinant for the choice under analysis. Indeed, the authors are here allowed to include seriously of illness, age and education of the individual, factors that plausibly influence expected health status.

Income is left to be included. The following equation describes the budget constraint faced by individual i:

$$C_{ij} + P_{ij} = Y_i \tag{2.4}$$

where Y_{ij} is individual income and P_{ij} is the monetary cost of provider j. Substituting for consumption equation 2.4 into Equation 2.1 yields:

$$U_{ij} = U(H_{ij}, Y_i - P_{ij}, T_{ij}) \tag{2.5}$$

In this formulation, income affects utility through the consumption term and is considered exogenous. The individual has $J + 1$ feasible alternatives (with $j = 0$ being self-care and 1 through J being medical care options). The unconditional maximization problem is thus given by:

$$U_i^* = max(U_{i0}, U_{i1}, ..., U_{iJ}) \tag{2.6}$$

where U^* is the maximum utility individual i can attain. The solution to Equation 2.6 reveals the alternative that is chosen. In a discrete choice model as the one under analysis, the demand function obtained can be interpreted as the probability that alternative j is chosen. An additional piece of information derivable from this formulation is the price elasticity of demand, which answers the crucial question of responsiveness of demand to a change in price of health service, or, in other terms, of the reaction to an introduction or removal of health user fees.

To conclude the theoretical background, a feature of health care as a good has to be underlined. Indeed, it is plausible to assume that health care is a normal good, whose demand increases as income increases. A necessary condition for normality, in this framework, is that the marginal rate of substitution of consumption for health decreases as income increases. In other terms, for high levels of income, individuals will be willing

to give up a higher amount of consumption for an additional unit of health[7]. This reasoning becomes particularly interesting in this framework, given that $Cij = Y_i - P_{ij}$: a low marginal rate of substitution of consumption for health implies that for high levels of income the individual is willing to pay a higher price (give up a higher amount of consumption) for seeking higher quality (for gaining an additional unit of health: note that health is expressed as $H_{ij} = Q_{ij} + H_{io}$)[8]. The reasoning can be reverted. Poor households will not be willing to give up a high quantity of consumption for an increase in one unit of health. This means, they will not be willing to pay a high price for having access to a high quality provider. Therefore, in light of these considerations, health care being a normal good implies price elasticity of demand for health care to be decreasing as income increases.

This theoretical background has been exploited for estimating the demand for health care through a discrete choice model. The following section outlines the econometric techniques adopted by the authors and describes how elasticity and welfare considerations are drawn.

2.2 Econometric Methods

In order to solve the problem stated in Section 2.1, it is necessary to attribute a functional form to the utility function, that will determine in turn the properties and characteristics of the demand function. Gertler et al. (1987) stress the importance of adopting a functional form that allows the marginal rate of substitution to be constant, increasing or decreasing as income rises. A linear functional form of the utility function would imply a constant marginal rate of substitution, imposing the restriction that income does not influence demand for health care and thus ruling out the possibility of health being a normal good. For this reason, the most commonly adopted is one of a utility which is quadratic in consumption and linear in health, which allows to test for the behavior of the marginal rate of substitution, without imposing it a priori. Let the conditional utility function be:

[7] And, implicitly, for seeking higher quality.
[8] This is proven graphically by Gertler et al. (1987).

$$U_{ij} = \alpha_0 H_{ij} + \alpha_1 (Y_i - P_{ij}) + \alpha_2 (Y_i - P_{ij})^2 + \alpha_3 T_{ij} + \epsilon_{ij} \tag{2.7}$$

with ϵ_{ij} a random shock uncorrelated across individuals and alternatives. Substituting Equation 2.3 into Equation 2.7 yields:

$$U_{ij} = \alpha_0 H_{i0} + \alpha_0 Q_{ij} + \alpha_1 (Y_i - P_{ij}) + \alpha_2 (Y_i - P_{ij})^2 + \alpha_3 T_{ij} + \epsilon_{ij} \tag{2.8}$$

Neither $\alpha_0 H_{i0}$ nor $\alpha_0 Q_{ij}$ are observed. Since the first term appears in all equations and it is not alternative variant, it can be ignored. Quality in itself is not directly observable, since as stated above Q_{ij} comprises both individual related element and provider related ones. Therefore, it can be approximated as follows:

$$\alpha_0 Q_{ij} = \beta_{0j} + \beta_{1j} X_i + \beta_{2j} Z_{ij} + \tau_{ij} \tag{2.9}$$

where X_i is a vector of individual characteristics such as age, seriousness of illness and education and Z_{ij} is a vector of provider j's characteristics faced by individual i. τ_{ij} is a random shock.

The utility function is obtained by substituting this last equation into Equation 2.8:

$$U_{ij} = \alpha_0 H_{i0} + \beta_{0j} + \beta_{1j} X_i + \beta_{2j} Z_{ij} + \tau_{ij} + \alpha_1 (Y_i - P_{ij}) + \alpha_2 (Y_i - P_{ij})^2 + \alpha_3 T_{ij} + \epsilon_{ij} \tag{2.10}$$

As easily deducible, the parameters of this utility function can not be estimated, since utility is in itself unobservable. However, as it is common in discrete choice models like this one[9], alternative k is chosen by individual i whenever $U_{ik} > U_{ij}$ for every $j \neq k$, and one can obtain the probability of choosing a provider j, once making distributional assumptions on the error terms.

Ching (1995)[10] assumes that the error terms follow a Type I extreme value distribution and this is known as a *conditional logit* estimation. The probability of choosing provider j is given by:

[9] For a theoretical summary and some practical applications see Cameron & Trivedi (2010).

[10] Gertler et al. (1987) adopt a nested logit specification, where first the individual chooses whether to seek care at all and then which kind of provider to have access to. Bolduc et al. (1996), in a study conducted in Bénin, compare different specifications (multinomial logit, multinomial probit and independent multinomial probit), showing how the results might turn out to be sensitive to the specification one decides to choose. For the sake of brevity, only the most widespread model is here analyzed.

$$\pi_{ik} = \frac{\exp\{V_{ik}\}}{\sum_{j=1}^{J+1} \exp\{V_{ij}\}} \tag{2.11}$$

where

$$V_{ik} = \alpha_0 H_{i0} + \beta_{0k} + \beta_{1k} X_i + \beta_{2k} Z_{ik} + \tau_{ik} + \alpha_1 (Y_i - P_{ik}) + \alpha_2 (Y_i - P_{ik})^2 + \alpha_3 T_{ik}$$

and

$$V_{ij} = \alpha_0 H_{i0} + \beta_{0j} + \beta_{1j} X_i + \beta_{2j} Z_{ij} + \tau_{ij} + \alpha_1 (Y_i - P_{ij}) + \alpha_2 (Y_i - P_{ij})^2 + \alpha_3 T_{ij}$$

Note that self-care is included in the J+1 alternatives. Equation 2.11 can be seen as a demand function for a provider k.

Authors have exploited price variations when a change of policy occurred in order to estimate the model above. The coefficients of interest at this stage are the ones attached to consumption and consumption squared, since their significance allows to establish whether price and income have some effect in the probability of choosing a certain provider. Both in Ching (1995) for the Philippines and in Gertler et al. (1987) for Perù the coefficients seem to be statistically significant[11], suggesting an impact of both income and price on the demand for health care[12].

The next and final step is the estimation of the price elasticity of demand. As underlined in Section 2.1, it is interesting to analyze the variation of elasticity according to the level of income, in order to assess whether data confirm the conjecture that health care is a normal good. What Gertler et al. (1987) and Ching (1995) implement is an estimation of price sensitivities of choice probabilities across five income groups. Both studies show how price elasticities are negative over all income groups and demand is more elastic at lower levels of income.

The model outlined above in a simplified version is an effective tool not only for em-

[11] At this stage, nothing can be said about the magnitude of the impact, since income and price enter the equation non-linearly.

[12] Different results are achieved by Akin et al. (1986) for the Philippines, same context examined by Ching (1995). The latter author claims this is due to an important theoretical difference: if this model allows for a natural interrelation between price and income, in Akin et al. (1986) these two elements are treated independently. However, in the Nigerian context examined by Akin et al. (1995), despite the similarity of the model to the one here exposed, once controlled for quality the magnitude of the effect of price on demand turns out to be relatively small and independent of income level.

pirical analysis, but also for a first-step reasoning about the issue in a quite comprehensive manner. Keeping in mind the limitations of any economic model, it is necessary to dig deeper into the analysis and investigate the different aspects highlighted in the debate about the positive or detrimental effects of health user fees. Just to mention a possible limitation of the above model, one can argue that it does not show any clear relation between an increase in price and an improvement in the service offered, as should be captured by the term Q_{ij}. Indeed, it is natural to believe that any fee generates revenues, which in turn should be the source of quality improvements. Additionally, this might encourage patients to seek more care than before, thus making a fee not detrimental for demand. To what extent an increase in quality occurs and how demand is consequently encouraged is an open empirical question, and has been topic for a considerable amount of research efforts. The following graph, proposed first by Mwabu & Mwangi (1986) and then by James et al. (2006) well summarizes the issue[13]:

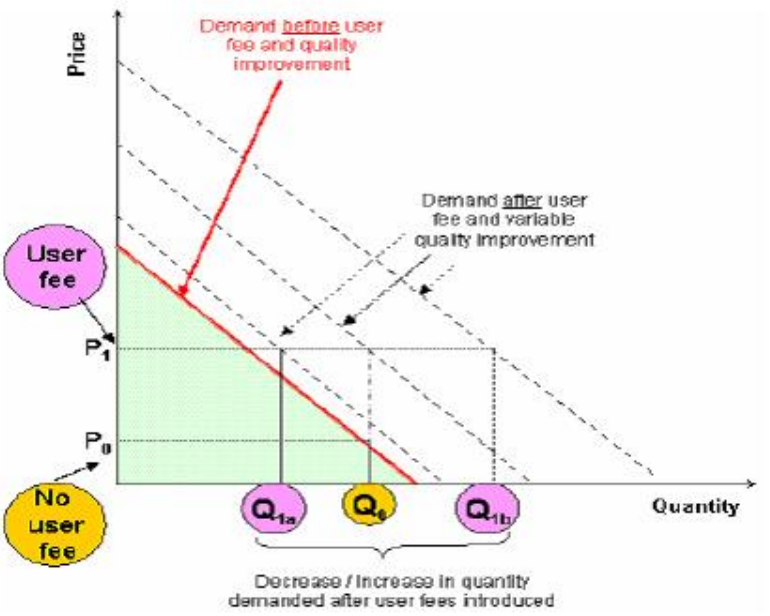


Figure 2.1: Health Care Demand. Source: James et al. (2006)

As can be seen, passing from P_0 to P_1 or vice versa, which means introducing or

[13]Interestingly, the core of the discussion seems to have been unchanged over twenty years. Nevertheless, as will be made clear in the following literature review, methodology, answers and policy implications have considerably varied over the period covered.

removing a fee, has an ambiguous effect on demand (on Q), depending on multiple factors: how large is the change in price? To what extent does the demand function shift because of quality improvements? Do these improvements occur at all? How elastic is the demand curve to price changes? Is this elasticity constant or does it vary according to variables such income, education and context? Authors have tried to address these topics, which are by their nature context-related and whose analysis is inevitably subject to methodological choices.

The next sections will propose some of the main arguments supporting the introduction of co-payments in developing countries and the ones supporting the opposite policy. The reader will gain an overview of the dynamics of the debate in 1980s and 1990s until the recent developments.

2.3 The 1980s and 1990s Debate

This section outlines the arguments that accompanied the introduction of health user fees in many developing countries starting from the end of 1980s. The stimulus for such a policy is clearly reconcilable to a document issued by the World Bank in 1987, whose main points deserve to be stated, in order to confront them with the empirical evidence.

2.3.1 An Agenda for Reform

One of the first points stressed by the international institution is the increased difficulty of governments in developing countries to accomplish increased public costs, among which and particularly the ones related to health care[14] (World Bank, 1987). Two are the causes of this increased burden: a general tendency towards slow economic growth and budget deficits in 1980s, on the one hand; an overall increase in health-related needs[15] together with urbanization, on the other hand. On this base, a new and alternative approach was proposed, having as first justification an important distinction.

Goods and supplied by the health system can be classified into two categories, private and public, according to the benefits they provide. The source of such difference can

[14]One of the reasons for the introduction of fees has been defined as *financial pragmatism* by James et al. (2006): central government expenditure was not reaching peripheral level health facilities, and thus user fee revenue could become the only source of finance for non-salary recurrent costs.

[15]The emergence of HIV/AIDS and its related health conditions, among these.

be traced back to who is the receiver of the beneficial effect, whether a single individual or the society as a whole. Generally speaking, in any context one can think of, there exist a wide spectrum of different goods, according to the nature of the benefit they provide. However, it is useful to define the two extreme categories: the one including purely private goods, whose benefits are enjoyed only by the person that purchases them, and the one into which fall purely public goods, for which all benefits are equally perceived by the society as a whole. In the health context, goods and services tend to fall into hybrid categories. The example provided by World Bank (1987) is the following: an aspirin for a headache has to be conceived as a purely private good. Instead, spraying to protect all residents from a vector-borne disease is a public good, since potentially everyone can equally benefit from it. Some hybrid categories include services like vaccinations, whose direct benefit is certainly perceived by the patient, but some indirect effect might positively influence the rest of the community, which will be less exposed to the illness: this phenomenon is known as positive externality.

What it is now important to underline is what economic theory suggests when it comes to explain consumer's behavior: agents are usually willing to pay for health services that guarantee some degree of private benefit. Instead, they will tend to be more reluctant in paying for services that equally benefit the community as a whole, behavior that is commonly known as the free-rider problem. The model thus proposed by the World Bank is the one exploiting this behavioral difference, by making individual paying for goods they are likely to be willing to pay (those who guarantee a solely private benefit, such as curative care) and letting the public intervene for goods or services that are likely to be exposed to the free-rider problem (such as immunizations, control of vector-borne diseases, health education, family planning).

Additionally, the document sheds some light on the source of problems for developing countries health care systems, listing three issues: an allocation problem, registered in an insufficient spending on cost effective health activities; an inefficiency problem, especially in public programs such as expenditure for drugs, fuel and maintenance; an equity problem, for the lack of balanced distribution of resources between high-cost hospital-based care and low-cost basic health services.

The World Bank proposes four specific policies for health financing, that should constitute parts of a package, reinforcing each other. Therefore, the introduction of co-payments is from the beginning thought of as a component into a wider range of com-

plementary government policies, as stressed one decade later by Gilson (1997). Charging users of government health facilities thus constitutes the first piece of advice. Particularly in developing countries, a design including differential fees for the poor should be implemented, being equal redistribution of resources one of the primary goals of the policy. Positive consequences of this practice could be a raise in revenues, an improvement on the quality of the service, financial sustainability of the system together with increased efficiency, provided by a reduction in frivolous utilization (unnecessary use) and in bypassing phenomena (Gilson 1997)[16]. Additionally, the introduction of health insurance schemes for risk sharing is recommended, together with policies for encouraging the efficiency of the non-government sector constituted by non-profit groups, community-run or privately managed cooperatives and traditional healers[17]. The last policy advice concerns the decentralization of planning and budgeting the services offering private benefits for which users are charged.

Following the World Bank policy advice, many developing countries started introducing co-payments adopting one of two broad models of user fee system (Nolan, 1995): the *traditional scheme* and the *Bamako Initiative scheme*. The former is in line with what recommended by the World Bank and recognizes in user fees a tool for raising revenues and produce resources, in turn producing efficiency benefits and protection for the poor, through differential charges. The system relies on price signals and incentives and it is implemented countrywide. This model is broad in nature and potentially adapts to any region in the developing world. What distinguishes the Bamako Initiative from this traditional scheme is the particular relation to the African context and the attention to the local dimension. Indeed, user financing should be managed at a community level in a participative manner, posing as a first aim the solution to the problem of under-resourcing of primary health care.

The framework described in this subsection should be of help for collocating the

[16]This point is stressed by Nolan (1995): a system of fees that reflects the relative cost of providing different types of services would encourage patients to ration their use of expensive resources and would prevent them from bypassing clinics or health centers and go straight to the high-cost hospital, as happens if both are free of charge. Later evidence, discussed in the following, prove different effects (Cohen & Dupas 2007).

[17]This category owns an important role in developing countries health practice, being usually well known at a community level and easily accessible, both from an economic and geographical viewpoint. These healers are characterized by a deep-rooted connection with traditional practices far from modern medicine and by lack of scientific knowledge, facts that tend to cause problematic consequences.

pieces of research that will be discussed in the following. First of all, space will be devoted to the analysis of empirical findings that confirm the arguments proposed by World Bank (1987)[18]. Afterwards, evidence on shortcomings and limitations of the policy will be highlighted, leading the discussion towards the recent change of direction, topic the chapter will conclude with.

2.3.2 Arguments in Favor of User Fees: Fees-for-Quality and Cost Recovery

One of the first contributions proving room for positive effects deriving from the introduction of user fees is provided by Akin et al. (1986), in a study that considers the Bicol region in the Philippines. By estimating the demand function for health care services in line with what exposed in Section 2.2, the author shows that once quality is accounted for, the economic costs of using medical care do not affect demand patterns, neither for services considered to be essential, like outpatient or delivery care, nor for services that are more optional in nature, like immunization or prenatal care. These findings shed some light on one of the main channels through which fees would lead to beneficial effects, namely in raising revenues and increasing quality.

A methodologically similar study conducted by Schwartz et al. (1988) confirms the previous results in another region of the Philippines, focusing on infant delivery systems. The authors carefully take into account quality factors such as the availability of modern public practitioners and facilities in rural areas, the amount of hours that health care facilities are open, availability of drugs, the presence of trained midwives for delivery, to show that the increase in these indicators would stimulate demand for health care. Moreover, the relative insensitivity of the choice of delivery service to changes in prices and income suggest a positive net effect of user fees on access.

Hotchkiss (1998) adopts a nested logit specification and makes s use of data collected both at a household level and at a facility level for investigating the identical issue, drawing again similar conclusions to the ones of the other studies above, despite showing a different result, supporting the presence of a negative price elasticity of demand for poor

[18]The same arguments are stressed by the World Bank in another document issued in 1995, which includes, additionally to what discussed in this section, some of the empirical evidence that will be reported below. See Shaw & Griffin (1995).

households. In simple words, what the studies cited so far suggest, is that in Figure 2.1 either the demand curve is very steep, thus not sensitive to price changes for all income groups, or the shift to the right due to an increase in quality following the rise in price is such that Q_0 tends to move to Q_{1b}.

What has to be underlined about these early studies is their purely theoretical and simulative character. Indeed, the results are not based on an actual evaluation of the impact of an implemented policy, but rather policy simulations designed to be informative to public officials interested in the effect of cost recovery schemes on utilization patterns. The claim for which introducing co-payments would stimulate demand for health care thanks to an increase in quality are based on cross sectional considerations and there is no poof of a potential raise in revenues and of a consequent quality improvement if such a policy was implemented.

An interesting study that overcomes the aforementioned limitation is the one by Litvack & Bodart (1993), who implement a pre-test post-test experiment in five health facilities in Cameroon, where a Bamako Initiative scheme was introduced[19]. The authors stress the importance of quality in any reliable analysis concerning such policies. Three health centers that were to introduce fees for quality improvements, especially reliable drug supply, were selected as treatment centers, whereas two facilities where such policy was still not into action were chosen as controls. Results from the strictly controlled experiments indicate that the probability of using the health center increased significantly for people in the treatment areas compared to those in the control areas. The channel through which this occurs contributes to confer another potential beneficial effect to user fees. Indeed, travel and time costs (that can be thought of as indirect costs of health care) involved in seeking alternative sources of care are high; when good quality drugs become available at the local health center, the fee charged for care and treatment represented an effective reduction in the price of care and thus utilization rose. Moreover, this study found that probability of the poorest quintile seeking care increases at a rate proportionately greater than for the rest of the population. Since the poor are most responsive to price changes, they appear to be benefiting from local availability of drugs more than others. One element has to be stressed: the cause of an increase in demand is again an increase in quality. However, this operates through a new channel, so far not yet

[19]As underlined in Subsection 2.3.1, the Bamako Initiative had a local community dimension, thus being particularly suitable for a controlled experiment.

considered. If an increase in quality was thought to make the demand curve shift, thus making agents willing to demand more for the same level of price, here the introduction of fees, the raise in revenue and the increase in quality[20] cause an overall reduction of the cost for health care, thanks to a cut of the indirect part of such cost. Analytically, borrowing some notation from Section 2.1:

$$TC_{ij} = P_{ij} + T_{ij} \tag{2.12}$$

where TC_{ij} is the total cost of health care, P_{ij} the direct component (fees) and T_{ij} the indirect component (transportation and time costs). The introduction of user fees ($P_{ij} \uparrow$) leads to an increase in quality, which in turn endows local health centers with drugs. This allows patients to access closer facilities, cutting transportation and time costs ($T_{ij} \downarrow$). Overall, if the reduction in T_{ij} outweighs the increase in P_{ij}, total cost decreases ($TC_{ij} \downarrow$) and demand increases, as it seems to have been the case in Cameroon.

Barber et al. (2004) propose and additional interesting aspect that could potentially fit to some developing countries contexts: the so called under-the-table payments. These are sort of informal fees, proven to constitute a high proportion of out-of-pocket payments in Cambodia, the environment under the authors' analysis. These fees, because *under-the-table*, are for their nature uncontrolled and can be interpreted as a signal of patients' willingness to pay. The results of the study, despite considering only a single health facility, show how introducing regular fees thus formalizing under-the-table payments can increase utilization. Indeed, the facility under analysis managed to control out-of-pocket patient expenditures, ensured patients of fixed prices, protected patients from the unpredictability of hospital fees and promoted financial sustainability. Not only utilization levels increased by more than 50% for inpatient and surgical services but cost recovery from user fees averaged 33%.

The studies reported above are only a representative sample of a wide set of other pieces of literature which have the same research question, are methodologically similar and are related to different countries and thus contexts in the developing world. Generally speaking, what can be inferred is that the efficiency of user charges is related to many factors: the level of externality, the price elasticity of demand in different income groups, the composition of total costs in terms of private access cost and opportunity cost (Mc-

[20] In the study by Litvack & Bodart (1993), this is related to endowing rural health centers with drugs

Pake 1993). If the previous studies show room for welfare gains for the majority, other research efforts have highlighted different mechanisms and effects, which are reported in the next subsection.

2.3.3 Arguments Against the Introduction of Health User Fees

Gertler et al. (1987), in the study mentioned for its theoretical framework, provides evidence for claiming that user fees are *regressive*, in the sense that user fees would reduce the access to care proportionally more for the poor than for the rich, being demand for health care more sensitive to price changes for low levels of income. The authors show that user fees can generate substantial revenues, but are accompanied also by important reductions in aggregate consumer welfare, with the burden of the loss on the poor. This result is at odds with what so far reported and similar conclusions have been drawn for a considerable number of contexts, additionally to the one of Peru in Gertler et al. (1987). A methodologically similar study, largely cited above, is the one by Ching (1995), where the claim of a regressive nature of user fees is supported for the Philippines, contrary to what other researchers conclude analyzing single regions, as in Subsection 2.3.2.

The consequences of co-payments for poor households have been largely investigated, and the majority of studies do not exclude a possible negative welfare effect for individuals with low level of income, despite a possible overall welfare gain for the society as a whole. Among these, interesting simulations are the one by Mwabu & Mwangi (1986) and Mwabu et al. (1993) for Kenya, where the typical price-quality trade-off phenomenon caused by user fees is observed and the overall welfare gain is opposed to equity considerations about the unbalanced character of such gain. Fabricant et al. (1999) analyze the context of Sierra Leone in 1990s, when the Bamako Initiative was implemented, about one decade earlier to the change in policy analyzed in Chapter 3. The authors, providing descriptive evidence for the regressiveness of health expenditure, insist on the importance of introducing differential co-payment schemes for poor households, especially for primary care services [21]. The survey conducted in the two regions of Kenema and Port Loko allowed to report the mean annual expenditure on treatment as a percentage of income by household income quintile. The poorest households spend one quarter of their income

[21] The authors recognize the difficulty in implementing such details, being economic status of households not only not recorded, but also hard to assess and measure.

for health care, a considerable fraction if compared to the 3% of households belonging to the highest income quintile [22]. A logistic regression allowed the authors to take into account different elements simultaneously, in order to assess the impact on choice of both economic and non-economic (distance, education, severity of illness and quality of the nearest health facility) factors. The results surprisingly show a quite inelastic demand to price for all income groups. Access factors such as walking time to a peripheral unit and distance to a hospital seem to play a more important role in the choice of medical treatment than most of the preference factors or the income and wealth variables. In this particular setting, the reason why poor and rural households are subject to a welfare loss is the following: in case of a severe illness, both rich and poor households would be willing to pay a certain amount of money for seeking care. However, since government facilities (those subject to the Bamako Initiative and thus providing proper health services) are too far from the village and distance plays an important role in the choice of provider, these households face a decision with only a limited range of alternatives, namely a very expensive one or non-formal medical treatment. For this reason, households already facing subsistence levels of income turn out to be dealing with a higher economic sacrifice for seeking care. Consequently, the study underlines how essential is the commitment of relating user fees to household ability to pay, since poor households are proven to be using the same facilities as richer households.

The heterogeneous reactions of different societies to a change in policy such as the one discussed so far makes the deduction of universal conclusions difficult and improper. There is universal agreement on what Lagarde & Palmer (2011) conclude in reviewing the literature on the topic: user fees constitute a barrier to health care, but can be source of quality improvements that in turn might have beneficial effects on the majority, even though in an unequal manner. Whether one tendency prevails on the other is a matter of time and space differences. So far, space differences have been the object of analysis, with attention given to different reactions to a simultaneous general policy change. The following section will move ahead in time, considering the New Millennium development in health care policy.

[22] The authors underline the higher probability of the poor of getting sick, among the possible causes.

2.4 The New Millennium Debate: Removal of User Fees for Universal Access

While in 1990s the literature mainly focused on assessing the validity of an introduction of user fees for increasing the quality of health facilities, thus addressing the beneficial supply-side effect of the issue, in the last decade increasing attention has been devoted on the positive demand-side effects of reducing user fees, not only in the developing world. The stimulus for this shift in research efforts and in policy decisions has come from the primary need of decreasing maternal and child mortality, as expressed in the Millennium Development Goals. To this end, encouraging access to health care through the abolition of fees for these sensitive categories has been considered to be an effective tool, in opposition to the described tendency of the 1990s.

With the development of new and more accurate methodological techniques for investigating developing countries economic and social mechanisms, more reliable evidence has been supporting the new tendency of abolishing co-payments for health services and deserves to be at least briefly mentioned. Holla & Kremer (2009) provide an exhaustive review of the evidence from randomized evaluations in developing countries on the impact of price on access not only to health care, but also to education. Two of these studies, being related to health outcomes, are of particular interest for this context. In the work by Kremer & Miguel (2004), an in-depth analysis about the illusion of sustainability in developing countries is conducted. As in many health cost-recovery programs listed above, the pursue of financial sustainability through fees seems to have revealed itself to be a mere and costly illusion, leading to undesired outcomes such as the one of a drop in health care utilization. A precisely designed randomized experiment for assessing the effects of a deworming program in Kenya was conducted. One of the most effective way of fighting intestinal worms, infecting one out of four people worldwide, is said to be periodical medical treatment with low-cost drugs. This treatment can be thought of as falling into the category of goods that provide both a private benefit and a public one (see Section 2.3.1), since roughly three-quarters of the overall benefit comes from the reduction of the disease transmission. The paper shows how the introduction of a small fee for deworming drugs led to an 80% reduction in treatment rates, consistent with the hypothesis that people have low private valuation for deworming. Take-up, essential in this kind of treatment, dropped sharply when going from a zero price to a positive

price but was not sensitive to the exact (positive) price level, suggesting that it may be particularly counter-productive to charge even small positive prices for the treatment of infectious diseases. This argument is strictly connected to the specific context of Kenya and to that particular deworming program, whose high positive externalities make it close to a public good, for which there might be lower willingness to pay. This, however, does not exclude the extendibility of the claim to other aspects of health care, especially in developing countries, where the awareness of the benefits of health care are somehow limited, making willingness to pay for it lower than one would expect. This is the reason why the claim of Kremer & Miguel (2004) can assume a wider validity, beyond the particular case of infectious diseases. As will be seen, this tendency of underestimating the beneficial effects of health care particularly suits the context of Sierra Leone analyzed in Chapter 3[23].

The second study confirming the just mentioned claims is the one by Cohen & Dupas (2007). The setting for a randomized malaria prevention experiment is again Kenya. The randomization concerns the price at which prenatal clinics could sell long lasting anti-malarial insecticide-treated nets (ITNs) to pregnant women. The argument for which user-charges would avoid wasting resources on those who will not use or do not need the product (in line to what expressed by the World Bank in 1987, as mentioned in Section 2.3.1) is here not supported: indeed, women who received free ITNs are not less likely to use them than those who paid subsidized positive prices for them. Additionally, the lack of cost for a net does not encourage frivolous utilization as in Gilson (1997), since those who pay higher prices appear not to be sicker than clients in the control group, receiving nets for free. Cost-sharing considerably dampens demand: uptake drops by 75% when the price of ITNs increases from zero to \$0.75. The authors conduct a cost-effectiveness analysis of ITN prices on child mortality that incorporates both private and social returns to ITN usage. Overall, given the large positive externality associated with widespread usage of insecticide-treated nets, the results suggest that in some settings free distribution might be as cost-effective as cost-sharing, if not more, exactly as in Kremer & Miguel (2004).

The straightforward question that arises when thinking of a policy that removes

[23]As a first element, one can take as an example the recent Ebola outbreak, considerably worsened by a lack of trust in the health care system, which encouraged affected people to seek care at traditional healers, thus facilitating the spread of the epidemic.

user fees concerns the way the health care system can be financially sustained, once co-payments are not anymore a source of revenue, and which are the possible drawbacks for the quality of the system as a whole. Indeed, as pointed out by Gilson & McIntyre (2005), as user fees restrict utilization of health services and create a large pool of unmet need, fee removal is likely to result in substantial and sustained increases in utilization[24]. Without increased funding for health care, these increases could well lead to falling quality of care generated by drug shortages and staff difficulties in managing increased workloads.

Many, in this respect, are the policy recommendations provided by the literature. Yates (2009) underlines the importance of governments and donors commitments in conceiving the financing of the health sector as a priority, expanding the budget for health care and increasing aid funds. The recommendation of Gilson & McIntyre (2005) is more specific in this respect. In order to reach universal coverage, African governments committed to allocate at least 15% of government budgets to the public health sector in the Abuja declaration of 2001. However, after the contribution of donors to health care funding is excluded, not many sub-Saharan African countries seem to have achieved this level so far[25]. For this reason, according to Gilson & McIntyre (2005), canceling African countries debt would support increases in tax revenues allocated to health services, by relieving government budgets from this burden. The author points out the possible limitations of the role of donor funding, because of its unreliable and discontinuous nature. Meessen (2009), in a UNICEF guidance note, stresses that fact that a government cannot just abolish user fees, but has to replace them with another way to fund the health system. Establishing this alternative financing option requires its quantification, the identification of its source, the long-term commitment of the sponsor, the institutionalization of the arrangement and the channeling of resources. Insurance schemes are seen as one of possible effective alternative sources of funding, as a form of risk sharing and a possible tool for reaching equal distribution of resources. This is the core message provided by Preker et al. (2005), in an extensive analysis related to how to spend wisely when providing health services for the poor. The lack of an acknowledged insurance system in the vast majority of developing countries makes this implementation hard, at least in the short

[24]Evidence supporting this is strong, and will be reported below.

[25]As reported by Pearson (2004), only Uganda and Mozambique exceeded the Abuja Declaration target of allocating at least 15% of Government spending to health. Not surprisingly, Uganda is one of the countries which is often brought as an example of a country where a well-structured and effective policy of removal of user fees was implemented, which will be described below.

run.

In a scoping review of the literature investigating the effects of the removal of health user fees in Africa, Ridde & Morestin (2011) provide an useful overview of the state of the art in investigating the effect of this widely implemented policy. Despite the need for further investigation, many studies provide evidence of mainly positive effects, especially in contexts where the initiative is accompanied by accurate implementation. Some of these studies will be hereby presented and particular attention will be devoted to Uganda and Ghana, examples of a well-design package of reforms.

Before getting into the analysis of single cases, a largely cited study proposing a simulation model for estimating the number of child deaths that might be prevented if user fees were removed in 20 African countries deserves to be mentioned, because focusing directly on a crucial health outcome, instead of on demand effects. In their study, James et al. (2005) implement a three-stage procedure for such a simulation. Firstly, a classification system for key intervention that might improve child survival is developed: 26 practices are classified according to whether their use is expected to increase after the removal of fees. Secondly, this intervention classification system is combined with evidence from some African countries on more generalized changes in use of health services after fee abolition, in order to produce estimates of expected changes in utilization rate for each of the 26 interventions. Finally, the authors convert estimated increases in use of different health interventions into plausible reductions in mortality for children under 5. The evidence provided by the study shows that abolition of user fees could have an immediate and substantial impact on child mortality, preventing an estimated 233,000 deaths annually in 20 African countries[26].

The pioneering country in the removal of user fees is the Post-Apartheid South Africa. In 1994, the new democratic government implemented a new reproductive policy that removed user fees from health care to pregnant women and children under 6 years old. The introduction of free health care led to substantial increases in access for black Africans, more than 70 percent of whom previously identified user fees as the major reason for forgoing treatment. Tanaka (2014) exploits the particular South-African context for implementing a difference-in-differences strategy, conferring to the study methodological

[26]Any generalized claim in this sense needs to be taken with caution, being many mechanisms at the source of such effects particularly context-specific.

credibility[27]. The main innovation proposed is to exploit plausibly exogenous variation in access to free health care among children with ex ante similar characteristics. Extreme domination by whites over the allocation of resources and limited mobility of black Africans under the apartheid regime created plausibly exogenous variation in availability of clinics across communities unrelated to the characteristics of black Africans. Therefore, households in communities with clinics were affected more by the abolition of user fees because they gained immediate access to health services, whereas those in communities without a clinic were less affected by the change, since they continued to have poor access to health clinics or had to wait until clinics were built. This heterogeneity enables the authors to apply a difference-in-differences strategy, which compares the changes in nutritional status between the communities with and without health clinics, before and after the policy change. Positive and significant effects are recorded for children's health status, captured by anthropometric indicators. Wilkinson et al. (2001), in an earlier study, show a sustained increase in new registrations and total attendances for curative services, but a fall in new registrations total attendances for services such as immunization and growth monitoring.

One of the countries in which removal of user fees is shown to have been particularly effective is Uganda. With the 2001 election campaign, President Museveni abolished cost sharing in the public sector, and fees were stopped in March 2001. To compensate for the loss of cost-sharing revenue and potential consequence on drug availability, the Ministry of Health introduced a supplemental buffer fund of 7 billion Uganda shillings ($5.5 million). This represented an increase of 22% to the ministry's drug budget for 2001 (Burnham et al. 2004).

Early evidence uniformly agrees in suggesting that improvements in utilization occurred and were most marked for the vulnerable category of poor people and women, as shown in Burnham et al. (2004), Nabyonga et al. (2005), Deininger & Mpuga (2004) Xu et al. (2006). Moreover, evidence from Deininger & Mpuga (2004) claims that the probability of getting sick decreased for the poorest categories. The big investment of the government for maintaining high levels of quality, threatened by the removal of an important source of revenue, seems to have been sufficiently effective: there was no de-

[27]As argued by the author, other studies among which the ones reported below for the context of Uganda tend to be before-after analyses, without cross-sectional variation. For this reason, the identification strategy of this study deserves to be outlined in detail.

terioration of indicators of quality such as cleanliness, compound maintenance and staff availability (Nabyonga et al. 2005). However, some aspects seem to not have been fully solved, as argued by Xu et al. (2006) in noticing how the incidence of catastrophic expenditures amongst the poor did not fall. According to the authors, this has to be attributed to the so-called drugs stock-outs, persistent in 2002 after the introduction of the policy despite fewer with respect to the years before, as pointed out by Nabyonga et al. (2005). This might have forced patients to purchase drugs from private pharmacies, with the extra payments for pharmaceuticals offsetting the reduction in payments for consultations. The challenge of sustaining quality persists, triggering sustainability of such a policy in the long run.

Very similar arguments were brought about by authors analyzing the Ghanian context[28], where a fee exemption policy for delivery care was introduced in September 2003. The way the policy was carried out allowed the analysis to be particularly sound, since the initiative was initially implemented in four most deprived regions of the country, Central, Northern, Upper West and Upper East. This allowed researchers to compare outcomes in one of these four treatment regions to outcomes of regions not affected by the policy. The comparison, in the vast majority of cases, occurs between Central (treated) and Volta (not treated), because similar in terms of socioeconomic characteristics.

Penfold et al. (2007) shed some light on the demand effects of the policy. As expected, the fee exemption increased the proportion of deliveries in health facilities, especially among the poorest and least educated women, the targeted group. Consequently, as shown by Asante et al. (2007), a statistically significant decrease in the mean out-of-pocket payments for cesarean section and normal delivery at health facilities was registered, but richer households benefited more from this merely economic viewpoint, having their expenditure decreased by roughly 20%, versus the 13% for the poorest households. Despite desired outcomes with respect to demand and households expenditure, Arthur (2012) shows how economic variables such as wealth keep on playing an important role in the access to health services, even after the implementation of a policy that should intuitively have equity effects, making access independent of economic status. With respect to health outcomes, Bosu et al. (2007) analyze the reaction of maternal mortality to the policy, comparing the aforementioned regions of Volta and Central. The delivery-related institutional maternal mortality ratio do not appear to have been significantly affected,

[28]Summarized by Witter et al. (2015).

at least after one year from the implementation of the policy, in the short run[29].

The extensive overview provided so far, concerning theoretical considerations together with a summary of the past and ongoing debate about user fees, will reveal itself to be necessary in order to interpret and place the evidence proposed for Sierra Leone in the following chapter.

[29]Consideration about short-run effects on health outcomes will be provided for the context of Sierra Leone in Chapter 3. As will be seen, results tend to be ambiguous as well.

Chapter 3

Evaluation of the Free Health Care Initiative

This chapter attempts to investigate the short-run effects of a change in policy occurred in Sierra Leone in 2010, the so-called Free Health Care Initiative (FHCI, hereafter). As widely argued in the previous chapter, since the beginning of the new millennium one of the common concerns expressed by governments of several countries in the developing world has been the reduction of financial barriers for accessing health care. This has been considered as a starting point for fulfilling important goals such as reducing mortality among sensitive categories like children and mothers. This introductory section will provide a detailed description of how the removal of health user fees was implemented in the country of interest, together with an overview about the structure and the characteristics of the health care system to which it was applied.

3.1 Institutional Setting and the FHCI (Free Health Care Initiative)

Sierra Leone is a West-African country among the 20 poorest in the world in terms of GDP per capita (World Bank, 2013). One of the main causes might be found in a long-lasting and devastating civil war that cost thousands of lives and negatively affected the economy and the already poor infrastructure. Sierra Leone is collocated in the low human

development category, with a Human Development Index of 0.374, ranked by the United Nations 183 out of 187 countries and territories in 2014 (Malik, 2014). In particular, according to the Sierra Leone DHS Final Report (2013) the maternal mortality ratio stood at 1,175 per 100,000 live births in the seven years preceding the survey and the child mortality rate at 155 per 1,000 live births. These indicators are still far from the 600 per 100,000 and 95 per 1,000 targets respectively of the UN Millennium Development Goals (MDGs) of 2015. Economic barriers to the access to health care have been considered a major cause, as proven by the low utilization of only 0.5 visits per person per year in 2009 (Ministry of Health and Sanitation, 2012).

In order to positively influence health indicators thus fulfilling the MDGs, the government aligned to the tendency of other poor African countries since the 2000s, encouraging higher utilization of health care by removing user fees for specifically targeted categories. At the United Nations General Assembly in September 2009 the President Ernest Bai Koroma announced that all health care services would have been made free for pregnant women, breastfeeding mothers and children under 5 years of age. The policy, which was first introduced on April 27th 2010, proposed the following measures: free care for pregnant women, breastfeeding mothers and children below 5 years of age in terms of free consultations including antenatal, postnatal and deliveries; treatment, minor surgeries, obstetric emergencies, basic emergency obstetrics and neonatal care (BEmONC) and comprehensive emergency obstetrics and neonatal care (CEmONC), x-rays and laboratory services, medicines and logistics (Ministry of Health and Sanitation, 2012).

The Free Health Care Initiative, embedded in a broader reform of the health system promoted by the Government of Sierra Leone (National Health Sector Strategic Plan, 2009), was applied to all public hospitals countrywide and was accompanied by some additional measures aimed at meeting the consequent increase in demand and at preventing a dramatic fall in quality of service, thus limiting the effect of the well known cost-quality trade-off. This includes the endowment of additional drugs and some general improvements of the facilities, with new machinery and additional health staff. These measures, trying to address the challenges of such a policy as underlined in Chapter 2, were not free of complications, not only as far as the collection of resources is concerned. Indeed, the logistics aimed at delivering the needed drugs were affected by phenomena

of corruption during the transportation phase[1]. Moreover, an initial adjustment phase was needed in order to fulfill the needs of health facilities to the increase in demand. In this framework, one of the most urgent and crucial measure to be implemented in order to make the FHCI effective at a broad level was the increase health workers number, availability and pay (Witter et al. 2015). This makes the Sierraleonian change of policy a catalyst for broader reforms of the health care system, recommended to reach the goal of sustainability and quality maintenance, as outlined in Section 2.4 in detail[2]. Moreover, as underlined by Witter et al. (2015), this constitutes a particularly precious opportunity for improving a health care system particularly fragile and conflict-affected such as the one of Sierra Leone[3].

The implementation of the policy was a result of a massive collaboration between the Government of Sierra Leone, the Ministry of Health and Sanitation, local authorities and development partners, crucial for facing an alternative financing scheme. The estimated cost to implement the initiative is of 35,840,173$, 86.5% of which provided by partners, mainly the Asian Development Bank, the Department for International Development (United Kingdom), United Nations Children's Fund, United Nations Population Fund and the World Bank.

Concerning the environment in which the policy was implemented, the organization of the public health system mirrors the administrative subdivision of the national territory. For each of the 14 districts, the so-called district hospital and the district health management team are the core of a network of peripheral health units (PHUs), which in turn are further sub-classified into three levels. Maternal and child health posts (MCHPs) are collocated at a village level for populations of less than 5,000 and are staffed by the so-called MCH Aides, trained for a limited range of services (antenatal care, supervised deliveries, postnatal care, family planning, growth monitoring and promotion for under-five children, immunization, health education, management of minor ailments, and referral of cases to the next level). Community health posts (CHPs) are at small towns level, for populations between 5,000 and 10,000. The staff is composed by state enrolled community health nurses (SECHNs) and MCH Aides, offering prevention and

[1]Not surprisingly, since corruption in the public sector is anecdotally known to be frequent in this country.

[2]Encouraging staff morale was one of the key elements.

[3]Unfortunately, the recent Ebola outbreak seems to have negatively affected it, as underlined in the recent paper by Witter et al. (2015).

control of communicable diseases and rehabilitation in addition to the services provided by the PHUs. More complicated cases are referred to the third level, the community health centers (CHCs) located at a chiefdom level and covering a population ranging from 10,000 and 20,000 (Ministry of Health and Sanitation, 2009).

Funding keeps on constituting a challenge for the health care system as whole. Public expenditure for the health system was only 8% of the total government expenditure in 2009, one year before the implementation of the FHCI (WHO, 2012), well below the 15% of the Abuja declaration of 2001. Despite a considerable increase in the expenditure in 2010, mainly caused by an increase in staff salaries in preparation for the change of policy, the funding of the health care system strongly relies on development partners, element that can potentially be seen as a challenge to the long term efficiency of the FHCI. The low amount of public spending is confirmed by the extremely high amount of private and out-of-pocket payments, which account for 69% of the total financing of the system and derives from spending on drugs by those who are not exempted by the policy. In addition to the target groups covered by the FHCI policy, some other groups are exempt from paying fees, such as emergencies cases, the destitute, and those with tuberculosis, HIV and other priority diseases. While there are funds to support the FHCI, there is no readily available funding for the other exempt groups. The health facilities are thus likely to face cash flow shortages, and this in turn can potentially affect the general performance of the FHCI. Moreover, there is no national health insurance system in Sierra Leone.

Fees for the non-exempted categories are quite high, considering that 57% live on less than 1$ a day and 74% on less than 2$ a day (World Bank, 2011). As an example, the Princess Christian Maternity Hospital in Freetown charges 5,000Le (1.15$) for consultation, 2,000Le (0.46$) for a follow up visit, 50,000Le (11.5$) weekly for admission and minor surgery, 200,000Le (46.03$) for a major surgery and 300,000Le (69.04$) for a complex major surgery.

There are a considerable amount of challenges to the beneficial effects of the FHCI. As mentioned, one of the primary concerns is related to the funding of the initiative and to a long-run sustainability of the project. Witter et al. (2015) conduct a study on how te FHCI has affected health workers. The authors highlight the positive effects of the policy, which triggered a series of reforms that significantly changed their number, pay and attendance. Moreover, despite an increase in the workload due to higher demand, the study provides evidence for an overall increase in reported motivation. However, tensions

remain between different cadres, since higher level staff seems to have benefited more than lower level staff and there is demand for a more consistent package of financial and non-financial incentives, particularly in rural areas.

Finally, anecdotal evidence suggests that many obstacles to the well functioning of the policy were caused by limited knowledge about the program for a vast majority particularly of the rural population. Moreover, phenomena of lagged or partial enforcement have been documented, with interviewed patients complaining of having to face unexpected out-of-pockets payments for special drugs or service, despite not being supposed to, as Amnesty International (2011) has reported. Indeed, in every facility there are two available types of drugs: FHCI drugs for three priority groups (pregnant and lactating women, children under-five) and cost recovery drugs (for all other categories), for which patients are charged. What researchers have noticed to occur is that when essential drugs are not available for women in pregnancy and childbirth for free, they are available for a price as cost recovery drugs. Additionally, delays in the delivery of drugs have been documented to be due to corruption and blockage, phenomena that are quite typical of this context.

3.2 Data and Descriptive Statistics

3.2.1 Data Sources

The source of data for this study is the 2013 Sierra Leone Demographic and Health Survey (SLDHS), which is the second population and health survey ever conducted in Sierra Leone. Data are collected every five years. All DHS surveys utilize a minimum of two questionnaires: a Household Questionnaire and a Women's Questionnaire. The main purpose of the Household Questionnaire is to provide the mechanism for identifying women eligible for an individual interview[4] and children under five who are to be weighed, measured, and tested for anemia. The dataset is thus organized in different recodes, each having a different unit of observation. This study exploits the birth recode, which has as unit of observation all live births of each eligible woman interviewed and includes useful information for the purposes of this analysis such as date of birth, place

[4]Women of reproductive age, that is between 15 and 49.

of delivery, anthropometric indicators, information on mortality, place of residence and GPS coordinates at a cluster level. Moreover, in order to estimate maternal mortality, the birth recode is merged with the individual recode, which has as unit of observation the eligible woman interviewed, in turn providing information about their female relatives' deaths linked to pregnancy or delivery[5].

Data are collected in 2013 using a two-stage stratified cluster sampling design. The sample is designed to produce estimates that are representative of the country as a whole, for urban and rural areas, for each of Sierra Leone four regions and 14 districts. This is the reason why stratification is applied[6] and within each stratum the sample selected in two stages. The first stage involves selecting primary sampling units (PSUs) called clusters, based on the list of enumeration areas (EAs) created in the 2004 Sierra Leone General Population and Housing Census. The enumeration areas provide the master frame for drawing 435 clusters (277 rural and 158 urban), selected with a probability proportional to their size[7]. In the second stage of the selection, 30 households were then systematically selected from each cluster. Following this methodology, a total of 13,006 households were selected for the sample, of which 12,724 revealed to be occupied. Among the occupied households, 12,629 were successfully interviewed, yielding a response rate of 99 percent. In the interviewed households, 17,132 eligible women were identified for individual interview; of these, complete interviews were conducted with 16,658 women, yielding a response rate of 97 percent.

Limitations of the dataset concern mainly the GPS coordinates, which I will use to compute distances to health facilities. Indeed, as described in depth in Section 3.4, they are available only at a cluster level (not at a household level), and they are subject to a random displacement. Moreover, for the purpose of the analysis, information related to health facilities such as number of cases, deliveries, vaccinations would be particularly useful in order to directly asses the change in demand for health services, as common in many similar studies. The dataset does not provide such information, which has to be indirectly made up by linking health outcomes and behaviors to the time period under

[5]The way maternal mortality follows the so-called sisterhood method, described in Section 3.4 and the main reference for the methodological aspect is Graham et al. (1989)

[6]Stratification is achieved by separating each district into urban and rural areas. Since the West Urban Area contains only urban areas, 27 sampling strata are constructed in total.

[7]The sampling frame excluded the population living in collective housing units, such as hotels, hospitals, work camps, prisons, or boarding schools.

investigation.

3.2.2 Descriptive Statistics

As will be explained in detail in Section 1.1, this study adopts a difference-in-differences strategy to investigate the effects of the FHCI for cohorts of mothers delivering (or children born) in 2010 comparing them to the ones for the cohorts of mothers (or children born) in previous years. The former cohort constitutes the treatment group for this study, whereas the latter the control group[8].

Table 3.1 provides baseline summary statistics of pre-FHCI variables for the treatment and control group respectively, comparing mean values between the two in order to test for significant differences. The period considered for the computation of statistics are the months from January to April, when the FHCI was still not implemented. *Panel A* anticipates which will be the outcome variables of interest, namely frequency of women delivering in a health facility, and some health outcomes like anthropometric indicators for children and mortality. As can be noticed by observing the last column, the two groups seem to differ significantly one another, as far as these variables are concerned. This might at a first sight suggest that these two groups are not proper for being compared. However, what one has to take into consideration is that the difference-in-difference strategy adopted for this study allows differences in levels for the outcome variables considered. The essential assumption is the so-called *common trend assumption*, which requires that the control group behaves the same way the treatment group would behave in the absence of treatment, namely following its same trend. This translates into the need for similarity of treatment and control group in terms of observable and unobservable characteristics, fact that is at a first stage confirmed by *Panel B*, reporting summary statistics for variables that this study will investigate in order to understand whether some categories are more responsive than others to the policy change. As can be seen, the two cohorts seem to be similar in terms of observable characteristics such as frequency of rich people, of educated respondents and of people living far from hospital (further than 30 km). Only the frequency of people living in rural areas differs across

[8]Justification for this choice will be provided in Section 1.1.

groups in a statistically significant manner [9].

3.3 Identification Strategy and Empirical Specification

The aim of this study is to analyze the short-run effects of the Free Health Care Initiative on several outcome variables: the frequency of women delivering in a health facility, maternal mortality, and health outcomes for children, namely anthropometric and mortality indicators. To this end, I will adopt a difference-in-differences strategy. In an ideal research environment, one would compare individuals that were exposed to the change of policy to others that could not benefit from it, both before and after the implementation of the policy. However, as mentioned in Section 3.1, the FCHI was put into action simultaneously countrywide in April 2010. This does not allow the identification of a control group as in Penfold et al. (2007), Bosu et al. (2007) or Asante et al. (2007) for Ghana, who compare regions where the policy was implemented first to regions similar in terms of socio-economic characteristics, but where the initiative would have been implemented later on[10].

For this reason, this study takes into account the cohort of women delivering (or children born) in 2010, the year when the FHCI was carried out. A naïve estimator of the treatment effect would simply compare the before-FHCI outcomes to the after-FHCI outcomes, within this cohort. One of the most important shortcomings of this kind of estimator is its inability to disentangle the effect of the policy from any other underlying element that might affect the outcome at hand, in any direction. In this short-run framework, one of the most relevant issues is the seasonality effect related to almost all of the outcome variables listed above. Both for phenomena such as reaching a health facility for delivering a child and mortality, seasonality effects play an important role in determining their trend. For this reason, an ideal control group encompasses these effects, that will then be taken into account and subtracted in order to isolate the treatment effect. Addi-

[9] Additional robustness checks will be shown in this respect in Section 3.5.

[10] One could argue, distance to a health facility could be a good source of variation for determining treatment status. However, as will be clear in the following, the data available are not sufficiently precise in this respect. Moreover, some additional anecdotal arguments related to common practices of moving and seeking hospitality in case of need to relatives that live closer to urban areas will be underlined.

Table 3.1: Differences at Baseline by Treatment Status

Variables	**Treatment Group**	**Control Group**	**Difference**
Panel A: Outcome Variables			
Frequency of women	0.51	0.37	0.13***
delivering in a health facility	[0.5]	[0.48]	(0.02)
Anthropometric Indicators:			
Weight for Height	-0.18	0.01	-0.19***
	[1.3]	[1.19]	(0.098)
Weight for Age	-1.12	-0.83	-0.29***
	[1.28]	[1.48]	(0.108)
Height for Age	-1.6	-1.4	- 0.2
	[1.77]	[1.79]	(0.139)
Maternal mortality Ratio	5927	3419	2507
(every 100,000 live births)	[3995]	[2076]	(2251)
Infant Mortality Rate	99.5	112.5	-12.9
(every 1,000 live births)			
Under Five Mortality Rate	127.5	149.4	-21.9
(every 1,000 live births)			
Panel B: Dimensions of control			
Frequency of people	0.71	0.66	0.05**
living in rural areas	[0.46]	[0.47]	(0.02)
Frequency of rich	0.57	0.58	-0.01
	[0.49]	[0.49]	(0.01)
Frequency of	0.819	0.816	0.003
low educated respondents	[0.39]	[0.39]	(0.01)
Frequency of	0.49	0.51	-0.02
people living far from hospital	[0.5]	[0.5]	(0.01)

Notes: The table provides summary statistics of pre-FHCI variables for treatment and control group. Panel A contains the outcome variables, whereas Panel B focuses on dimensions of control. The last column estimates the differences in means between the two groups under the null hypothesis that the difference is equal to zero. Standard deviations are in square brackets and standard errors are in parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

tionally, one can argue that apart from the seasonality effects, there might be a long-run underlying trend to every outcome variable here considered. I address these two issues by building a simulation of the counterfactual as follows: the control group is composed by cohorts of the six preceding years, whose outcomes are averaged, allowing to embody the mentioned patterns of seasonality.

The identifying assumption is that the average performance across these cohorts simulates the behavior of the 2010-cohort in the absence of treatment. In other words, the control group embodies both patterns of seasonality and the long-run underlying trend, which is assumed to have a constant pace (absence of acceleration). Stated differently, the assumption is of a similarity in terms of observable and unobservable characteristics of the two groups, as already partially confirmed by the descriptive statistics of Panel B in Table 3.1. As for the anthropometric indicators for children, one can reasonably assume that seasonality effects are not relevant: a child born in January is not more or less likely to be healthier than a child born in July [11]. For this reason, only the cohort of children born in 2009 is considered as control group.

One can easily reconcile this to the most common difference-in-difference framework, where two sources of variation are taken into account: a cross-sectional variation (treatment and control) and a before-after dimension. One should not get confused by the fact that in this study the one that should be a cross-sectional variation has somehow a temporal connotation, since cohorts delivering or born in preceding years are taken into account[12]. For the reasons explained above, one can reasonably conceive this control group as the counterfactual, or, in other terms, a simulation of how the treatment group would behave if not treated. The before-after dimension of the analysis is preserved: outcomes of both groups are compared before and after the FHCI. To summarize, following the notation of Section 1.3, the treatment effect is defined as:

$$D_{Did} = [\hat{\mathrm{E}}(Y_1|T) - \hat{\mathrm{E}}(Y_0|T)] - [\hat{\mathrm{E}}(Y_1|C) - \hat{\mathrm{E}}(Y_0|C)] \tag{3.1}$$

[11] As instead this reasoning can apply to the other outcomes, for quite straightforward reasons.

[12] When it comes to consider anthropometric indicators, one can reason in another way, considering the idea that different age cohorts of children are taken into account. This point will be stressed again later when discussing the results.

or, equivalently:

$$D_{Did} = [\hat{\mathrm{E}}(Y_1|T) - \hat{\mathrm{E}}(Y_1|C)] - [\hat{\mathrm{E}}(Y_0|T) - \hat{\mathrm{E}}(Y_0|C)] \tag{3.2}$$

In a regression framework:

$$Y_{it} = \alpha + \beta \cdot 1(t = 1) + \gamma \cdot 1(i \in T) + \eta \cdot 1(i \in T)1(t = 1) + \epsilon_{it} \tag{3.3}$$

where η is the parameter of interest, $1(t = 1)$ is a dummy for the period after the implementation of the policy, $1(i \in T)$ is a dummy for treatment and ϵ a normally distributed error term.

The identification strategy is similar for every outcome variable. Moreover, I analyze the sensitivity of different categories to the FHCI, repeating the previous regression separately for rural and urban mothers, rich and poor, close and far from a health facility, educated and non-educated. The following section goes in depth into the analysis of every single outcome and reports the empirical results.

3.4 Empirical Results

3.4.1 Demand: Delivering in a Health Care Facility

In the vast majority of studies mentioned in Chapter 2, one of the most investigated issues both for the introduction and for the removal of user fees is the reaction of demand. In the setting analyzed by this chapter, one of the ways to assess whether the FHCI increased demand for health care is to observe the frequency of mothers deciding to deliver in a health facility, being them the target group, together with children under five. The aim of this section is to verify whether the policy had some effect on the frequency of women enjoying skilled assistance while delivering, in a country where delivering at home unaided or supported by traditional birth attendants (TBAs) is still a very common practice.

Figure 3.1 shows the frequency of women delivering in a governmental health facility[13], computed for groups of two months both for the treatment group and for the

[13]Women delivering in private clinics are not considered, since the policy concerns the removal of user fees only in public facilities.

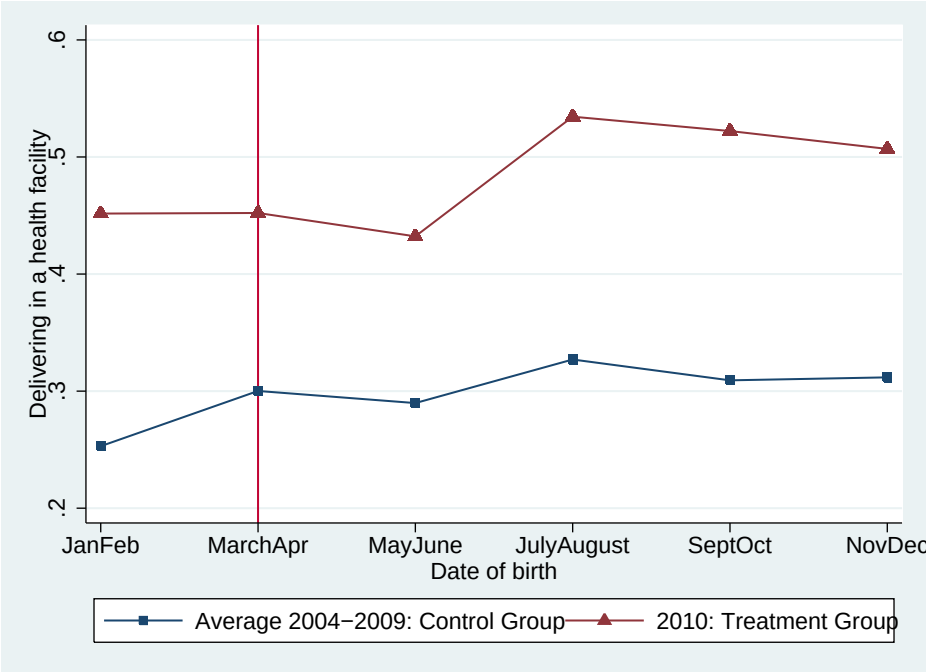


Figure 3.1: Delivering in a Governmental Health Facility: Trends

control group, where for each group of months the average frequency is computed for the six years preceding the policy. As said, this makes it embody patterns of seasonality and an underlying trend, thus mirroring how the treatment group would have behaved in the absence of the program. One can easily notice how in terms of levels, treatment and control group differ significantly one another (see Table 3.1 for the baseline levels). With respect to the trend for the treatment group, the figure suggests that there has been a considerable increase in the demand for health care by women having to deliver their child. Moreover, it shows how the reaction has been somehow lagged: the vertical red bar indicates the moment at which the policy was implemented, but the major jump is evident only after June, approximately two months later. It does not come as a surprise, since it is very plausible to assume a delay in the spread of information on the one hand and in the full implementation of the policy on the other, especially in the aspects related to endowing hospitals with additional equipment and drugs[14].

Table 3.2 reports the difference-in-difference estimation of the treatment effect, considering as post-FHCI period the one from July to December, for the just mentioned

[14]As mentioned in Section 3.1, phenomena of corruption and delays have been documented.

Table 3.2: Difference-in-differences Matrix: Delivering in a Governmental Health Facility

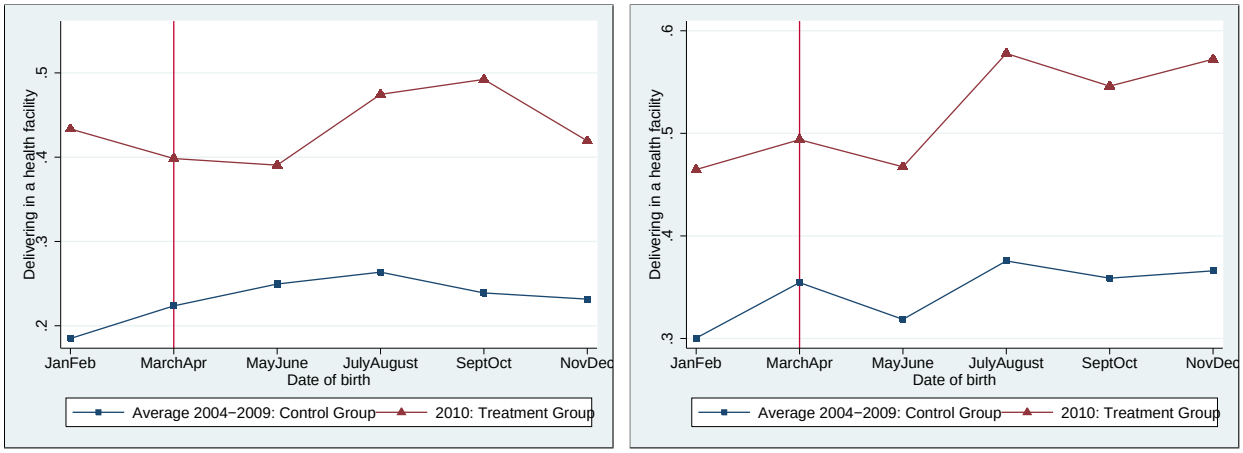
	Frequency of women delivering in a health facility		
	Treatment Group	*Control Group*	**Difference**
Post-FHCI (July-December)	0.522	0.316	0.206***
	(0.021)	(0.012)	(0.021)
Pre-FHCI (January-June)	0.445	0.282	0.163***
	(0.020)	(0.011)	(0.020)
Difference	0.077**	0.034**	**0.044***
	(0.013)	(0.016)	(0.024)

Notes: Each entry in cell reports the mean value of the frequency of women delivering in a health facility (computed for each single month in the year as the number of women going to hospital over the total of mothers delivering) in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Clustered standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.



(a) Poor (b) Rich

Figure 3.2: Delivering in a Governmental Health Facility: Rich vs Poor

reasons. Each cell reports the average frequency for the period and group considered, respectively. The table suggests that thanks to the FHCI alone, the frequency of women delivering in a health facility increased by 0.044, effect that is significant at the 10 percent level. In other words, the rate of utilization increased by 4.4 percentage points. The effect evidenced here is a short-run effect, and even if in Figure 3.1 the effect seems to be sustained until the end of the year, nothing can be inferred for the long run.

What one can conclude at a first stage is that the FHCI has encouraged access to health facilities, thus increasing demand for health care: similar conclusions to the ones encountered in studies relative to other countries.

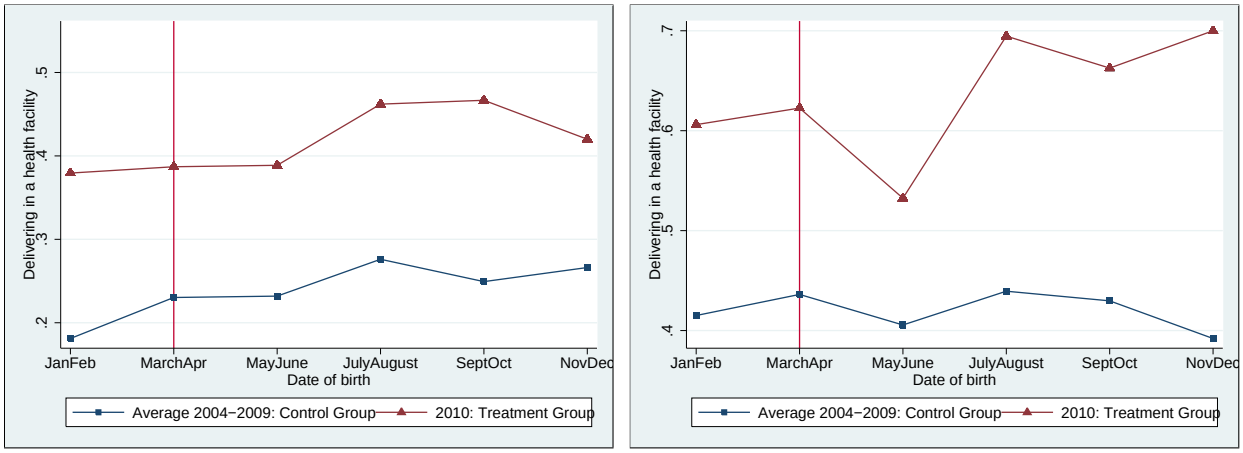
It is interesting to verify whether the initiative had some heterogeneous effects with respect to certain dimensions. For instance, it is essential to investigate whether poor mothers are more responsive to the change in price, as one would expect, under the commonly made assumption that the elasticity of demand is higher for poor individuals (as discussed and supported with empirical evidence in Chapter 2). To this end, I repeat the analysis separately for poor mothers, on the one hand, and for rich mothers, on the other. I consider *poor* the ones that belong to the first and second income quintile, *rich* the others[15].

[15]In the DHS dataset, income quintiles are defined according to a wealth index including variables such as ownership of a house, presence of electricity, source of drinking water, location of the kitchen, ownership of a car.

Figure 3.2 compares trends for the treatment and control group separately for poor and rich women. As can directly be seen from the graphs, the increase in the frequency seems to be more pronounced within the group of rich women (again, only after June), contrary to what one would expect. As for the poor women, there seem to be a positive effect as well, even though not sustained until the end of the year, thus leaving open the question of whether it is due to seasonality effects or to an actual decrease in the effect. Panel A of Table 3.3 reports the difference-indifference matrix for the poor, showing a low and not statistically significant effect. The effect for rich women is instead almost double and statistically significant. This is at odds with the theory that supports a higher elasticity of demand for poor households. There are at least two explanations for this phenomenon. First of all, one can plausibly assume that poor households are the least educated, in a country where school fees are an important determinant of educational choices. As will be seen, the policy had a major impact on more educated mothers and consequently, if the reasoning is true, on the richest. A similar argument can be made by assuming that poor households are also the ones that live in rural areas[16] (as will be seen, the effect for rural individuals is not significant). A second interpretation can be made with respect to richer mothers: once removed user fees and increased the quality of governmental health centers, it is possible that both women that would not have gone to hospital and women that would have chosen private health facilities shift to public ones, thus leading to a high and significant effect.

Another interesting dimension to explore is the responsiveness of households living in rural areas in comparison to the ones living in urban areas. The distribution of the effect is hard to predict ex-ante. Indeed, one could argue that the highest effect should be registered in rural areas, where people were probably less keen on going to hospital both for distance reasons and for economic reasons, if one assumes that rural households are poorer. On the other hand, the higher presence of health facilities in urban areas or higher opportunities to reach the health centers thanks to public transport might make households living in the city more responsive to the policy change. The latter theory is the one confirmed by data, and is again evident only by observing and comparing the trends in Figure 3.3. The effect for urban women looks particularly pronounced and this is confirmed by Table 3.4: the frequency significantly increased by more than 9 percentage

[16]I test for this by controlling for education and rural residence: there is no significant difference between the effect for poor and rich, at that point. See Appendix.



(a) Rural (b) Urban

Figure 3.3: Delivering in a Governmental Health Facility: Rural vs Urban

points, versus the almost not significant 2 percentage points for rural women. One of the possible explanation to this difference in responsiveness can be identified in the big role of traditional birth attendants in rural areas and, as said, a lower diffusion of governmental health facilities.

Partially related to what has just been discussed, another crucial dimension to investigate is the role of geographic distance to a health facility when it comes to decide whether to seek health care in a health facility or to find alternative sources of care. As more than once highlighted in Chapter 2, the total cost of health care for a household is a combination of both direct economic costs (such as money spent for user fees and transportation) and indirect costs such as traveling time. Before reporting the results for close and far households, it is necessary to highlight some methodological issue, related to how distance is computed.

The DHS dataset does not include any variable directly reporting distance to the closest health facility. For this reason, I exploit the GPS coordinates of each cluster available in the DHS dataset and compute the distance to the closest health facility. I use a list of 62 facilities, including a total of 43 hospitals and a selection of the major community health centers and health posts. The source of the list of health facilities and the respective GPS coordinates is a Crowd Sourced Hospital Information Dataset[17], where additionally to the GIS data, a list of the main health care services offered is pro-

[17]Knoema.com, 2014-2015

Table 3.3: Difference-in-differences Matrix: Delivering in a Governmental Health Facility (Rich vs Poor)

	Frequency of women delivering in a health facility		
	Treatment Group	*Control Group*	**Difference**
Panel A: Poor			
Post-FHCI (July-December)	0.460	0.245	0.215***
	(0.031)	(0.016)	(0.032)
Pre-FHCI (January-June)	0.406	0.219	0.187***
	(0.029)	(0.015)	(0.029)
Difference	0.054**	0.026**	**0.029**
	(0.020)	(0.02)	(0.037)
Panel B: Rich			
Post-FHCI (July-December)	0.565	0.367	0.199***
	(0.025)	(0.015)	(0.025)
Pre-FHCI (January-June)	0.475	0.326	0.148***
	(0.024)	(0.013)	(0.024)
Difference	0.090***	0.041**	**0.050***
	(0.022)	(0.018)	(0.030)

Notes: Each entry in cell reports the mean value of the frequency of women delivering in a health facility (computed for each single month in the year as the number of women going to hospital over the total of mothers delivering) in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Clustered standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 3.4: Difference-in-differences Matrix: Delivering in a Governmental Health Facility (Rural vs Urban)

	Frequency of women delivering in a health facility		
	Treatment Group	*Control Group*	**Difference**
Panel A: Rural			
Post-FHCI (July-December)	0.450	0.264	0.187***
	(0.027)	(0.014)	(0.026)
Pre-FHCI (January-June)	0.385	0.215	0.170***
	(0.024)	(0.012)	(0.023)
Difference	0.066***	0.049**	**0.017**
	(0.012)	(0.019)	(0.028)
Panel B: Urban			
Post-FHCI (July-December)	0.685	0.422	0.263 ***
	(0.028)	(0.021)	(0.032)
Pre-FHCI (January-June)	0.588	0.420	0.167
	(0.030)	(0.018)	(0.032)
Difference	0.098***	0.002	**0.096****
	(0.029)	(0.026)	(0.040)

Notes: Each entry in cell reports the mean value of the frequency of women delivering in a health facility (computed for each single month in the year as the number of women going to hospital over the total of mothers delivering) in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Clustered standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

vided. This allows to exclude facilities that do not supply any maternal and/or pediatric service. After having filtered the list in light of this additional information, I compute the number of kilometers to the closest health facility using geodetic distances.

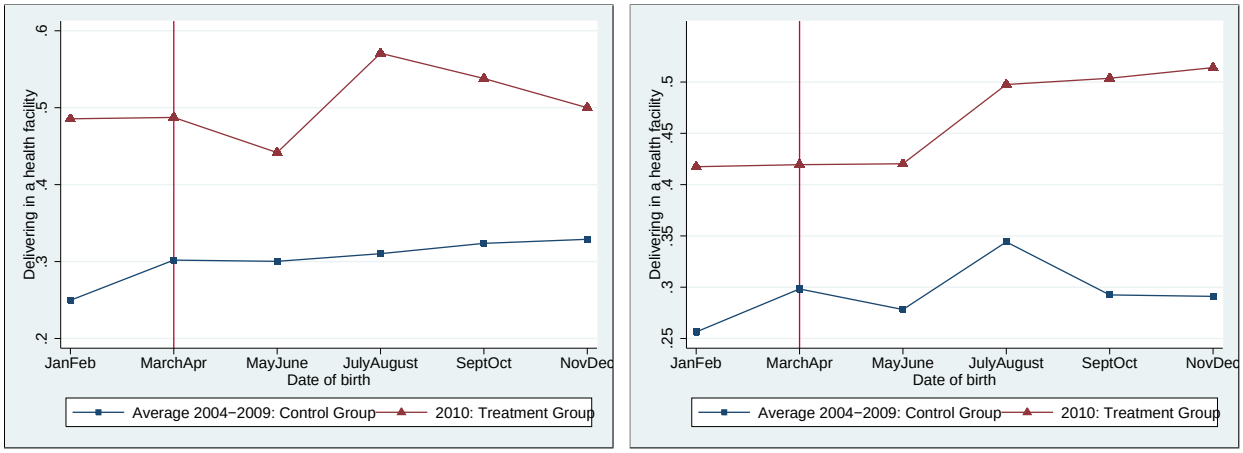
While the coordinates for hospitals are reliable, some limitation is instead identifiable in the GPS coordinates for households. First of all, DHS provides GPS data only at a cluster level and not at the household level. Moreover, for privacy reasons, the real GIS data are randomly displaced, that is, a random distance is applied (within 1km for urban clusters, up to 5km for rural clusters). This leads to some bias in the computation of geodetic distance. For this reason I implement a quality-test for this variable comparing it to a self-reported measure of distance included in the survey. The respondent is asked to express an opinion regarding the relevance of distance to a health center when getting medical help for self and three categories emerge: those for which it is a big problem, not a big problem, or no problem. The average distance for the first group shows to be 32km, for the second 25km and for the third 22km. This sheds some light on the reliability of the computed variable, despite the likely presence of a bias. Consistently with the quality-test, I choose 30km as a threshold for classifying close or far households.

Figure 3.4 reports the trends of the frequency of delivering in a health facility for close and far women. By observing the trend for the control group in both categories (close/far), one can easily notice how the frequency level is quite similar, contrary to what one would assume. This might be due to the lack of precision of the distance measure, as underlined above[18]. Despite this, under the assumption that the trend for the control group embodies seasonality effects, one can easily notice how these are more pronounced in the group of mothers that are farther away, and this looks as one would expect. As for the effect of the FHCI on the rate of utilization of governmental health facilities for delivering, the result in this respect might look surprising at a first sight. Indeed, as shown by Table 3.5, the effect on closer mothers looks considerably lower than for mothers that live more than 30km away from a health facility[19].

A possible explanation for this can be given by assuming that the economic barrier is one of the most relevant when having access to health care. Once this barrier is eliminated, even mothers that live far away seem to be willing to have access to health

[18]When using self-reported measures of distance, the frequency of close women delivering in a health facility is overall higher than for the far ones (both for the treatment and for the control). See Appendix.

[19]Results using the self-reported measure of distance are very similar, as shown in the Appendix.



(a) Close (b) Far

Figure 3.4: Delivering in a Governmental Health Facility: Close vs Far

care, maybe seeking hospitality by relatives that live closer to a hospital in the weeks preceding the delivery. In other words, the policy might have encouraged far households to commit more than before for reaching the fee-free hospital. There is some anecdotal evidence that supports this view. Some English volunteers operating in Freetown (the capital city) claim: *In order to seek the necessary treatment, [households] would uproot their families and make the incredibly difficult journey across the country to reach the capital, bypassing their regional health care centers in the process*[20].
This and the results reported in Table 3.5 might suggest that households living very far away from hospitals, once aware of the change of policy, show to be willing to travel a lot in order to reach the facility, conscious of the fact that transportation and time costs will be counterbalanced by zero direct costs.

Finally, the last dimension here investigated concerns the impact of education on encouraging mothers to reach a government health care facility after fees were removed. Mothers are categorized with *Low Education* if they attended less than 5 years of school and with *High Education* otherwise. Figure 3.5 reports the trends, whereas Table 3.6 the difference-in-differences matrix. Highly educated mothers are more responsive to the policy change, with a statistically significant increase in the frequency of 0.044 versus the not significant 0.038 of less educated women. This suggests that education plays an important role when it comes to health choices. Consequently, removing economic

[20]ICS (International Citizen Service), 2013

Table 3.5: Difference-in-differences Matrix: Delivering in a Governmental Health Facility (Close vs Far)

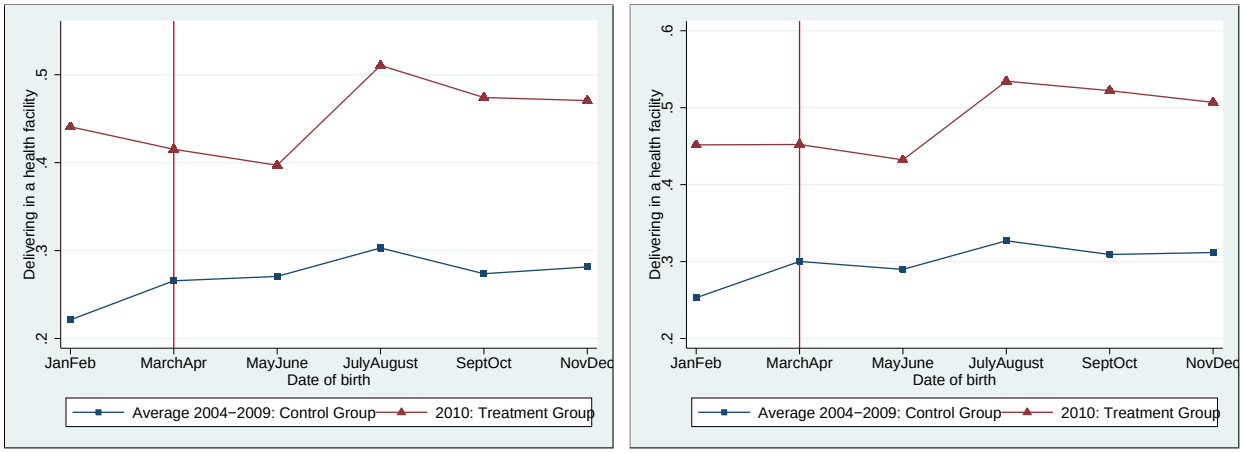
	Frequency of women delivering in a health facility		
	Treatment Group	*Control Group*	**Difference**
Panel A: Close (⩽30km)			
Post-FHCI (July-December)	0.539	0.321	0.218***
	(0.027)	(0.017)	(0.028)
Pre-FHCI (January-June)	0.471	0.286	0.186***
	(0.026)	(0.015)	(0.026)
Difference	0.068**	0.035	**0.033**
	(0.023)	(0.020)	(0.033)
Panel B: Far (>30km)			
Post-FHCI (July-December)	0.506	0.309	0.197***
	(0.007)	(0.014)	(0.015)
Pre-FHCI (January-June)	0.418	0.278	0.140***
	(0.009)	(0.011)	(0.015)
Difference	0.088***	0.031	**0.057***
	(0.012)	(0.017)	(0.021)

Notes: Each entry in cell reports the mean value of the frequency of women delivering in a health facility (computed for each single month in the year as the number of women going to hospital over the total of mothers delivering) in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Clustered standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.



(a) Low Education (b) High Education

Figure 3.5: Delivering in a Governmental Health Facility: Low vs High Education

barriers for encouraging access to health care is not enough and needs to be associated to other measures in the field of education.

This first subsection has shown the impact of the FHCI on the demand for health care, measured in this context by the frequency of women delivering in a governmental health facility. The following subsection tries to assess whether the policy had some short-run effects on health outcomes for children under five.

3.4.2 Health Outcomes: Anthropometric Indicators

The aim of this subsection is to assess whether the FHCI had some beneficial effect in improving children health status, beyond a mere increase in demand. As said, the initiative provided free care for pregnant and lactating women and children under five years of age. Therefore, the channels through which children might have benefited are two: antenatal care and postnatal care, both of which after the implementation of the policy became free of charge.

Anthropometric indicators such as weight for height, weight for age and height for age are a useful tool in order to test for children health status[21]. Indeed, they are a good summary both of nutritional aspects (in this case mainly related to breastfeeding) and of other aspects such as overall health conditions. The identification strategy is

[21] They are z-scores, measured in terms of standard deviations with respect to an ideal point of reference. See Appendix for a short paragraph on anthropometric indicators.

Table 3.6: Difference-in-differences Matrix: Delivering in a Governmental Health Facility (Low vs High Education)

	Frequency of women delivering in a health facility		
	Treatment Group	*Control Group*	**Difference**
Panel A: Low Education ($\leqslant$ 5 years)			
Post-FHCI (July-December)	0.485	0.285	0.199***
	(0.024)	(0.013)	(0.023)
Pre-FHCI (January-June)	0.416	0.254	0.162***
	(0.021)	(0.012)	(0.021)
Difference	0.069***	0.031	**0.038**
	(0.019)	(0.018)	(0.027)
Panel B: High Education ($>$ 5 years)			
Post-FHCI (July-December)	0.522	0.316	0.206 ***
	(0.021)	(0.012)	(0.021)
Pre-FHCI (January-June)	0.445	0.282	0.163***
	(0.020)	(0.011)	(0.020)
Difference	0.077***	0.0334**	**0.044***
	(0.013)	(0.016)	(0.024)

Notes: Each entry in cell reports the mean value of the frequency of women delivering in a health facility (computed for each single month in the year as the number of women going to hospital over the total of mothers delivering) in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Clustered standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

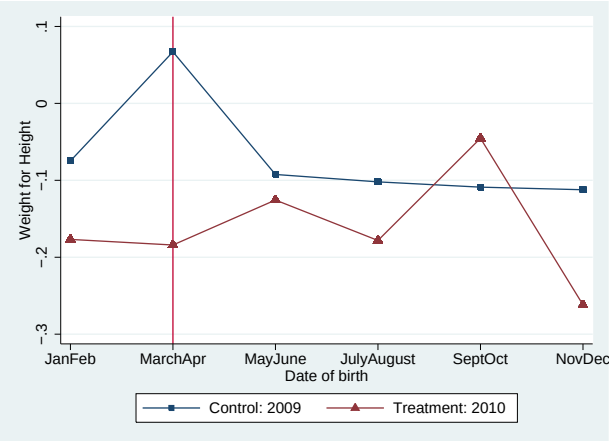
similar to the one implemented for the effects on demand, with the only difference that the control group here only includes the cohort of children born in 2009, assuming that such indicators are not affected by seasonality effects and one is instead interested to keep track of the underlying trend not related to the policy change, which constitutes the counterfactual.

Figure 3.6 reports the trends for the aforementioned groups for three indicators. Only weight for height seems not to have been affected by the policy. Instead, directly after the implementation of the initiative, both weight for age and height for age seem to have improved considerably. Table 3.7 reports the difference-in-differences matrices and shows how only the effect on weight for age is statistically significant (the effect on height for age is higher, but not significant). The point estimates suggest that free health services increased average weight-for-age z-scores by 0.27 standard deviations. An important clarification has to be done. For the estimation of the effect I use the DHS dataset for 2013. This means that the cohort of children both born in 2009 and in 2010 are measured in 2013, 3 years after the policy was implemented. For this reason, children in the control group are likely to have been partially exposed to the policy, but only after they reached 1 year of age. The effect reported in Table 3.7 is thus likely to underestimate the overall effect of the policy. However, this allows to isolate the effect on children of only early care[22] and of a plausibly healthier mother breastfeeding them. Despite these aspects, the identification of a positive and significant effect sheds light on the importance of these sources of care for newborns, additionally to plausibly healthier mothers.

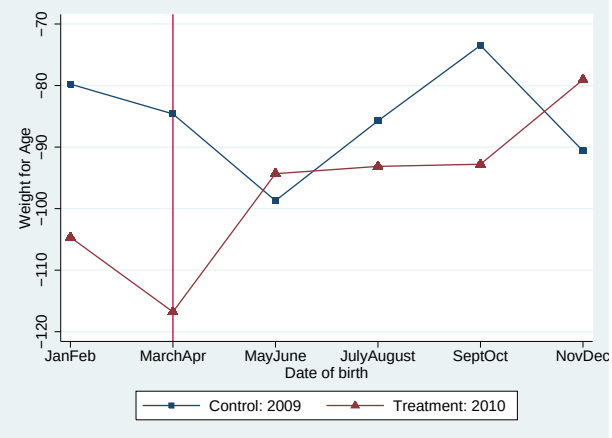
The analysis of the heterogeneous responsiveness of specific categories, as conducted in Section 3.4.1, will only refer to weight for age[23]. As before, I investigate whether the effect varies across income groups. Figure 3.7 reports trends, which show a general improvement for both groups. Moreover, one can notice how poor children are overall worse off compared to the rich ones, as it is after all expected. Contrary to what occurred for the frequency of women delivering in a health facility, the effect is considerably higher and statistically significant for poor children, as evidenced by Table 3.8. The policy

[22]Note another aspect that might underestimate the effect: children born in the months directly after the implementation have not benefited from free antenatal care: the effect is only related to postnatal care.

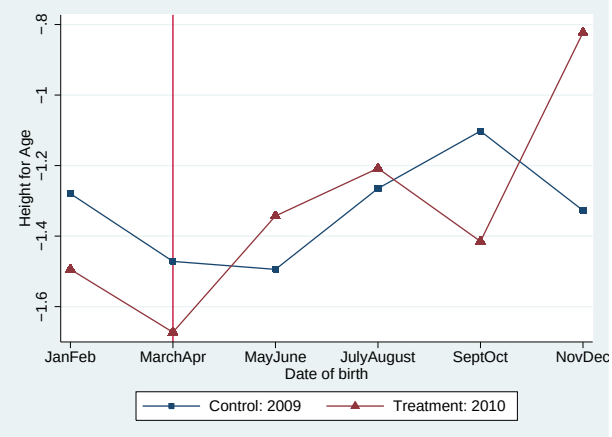
[23]The same choice is made by Tanaka (2014) for the analysis of the South African context, where, as here, only this short-run indicator of nutrition and health showed to have been positively influenced by a similar policy.



(a) Weight for Height



(b) Weight for Age



(c) Height for Age

Figure 3.6: Anthropometric Indicators: Trends

Table 3.7: Difference-in-differences Matrix: Anthropometric Indicators

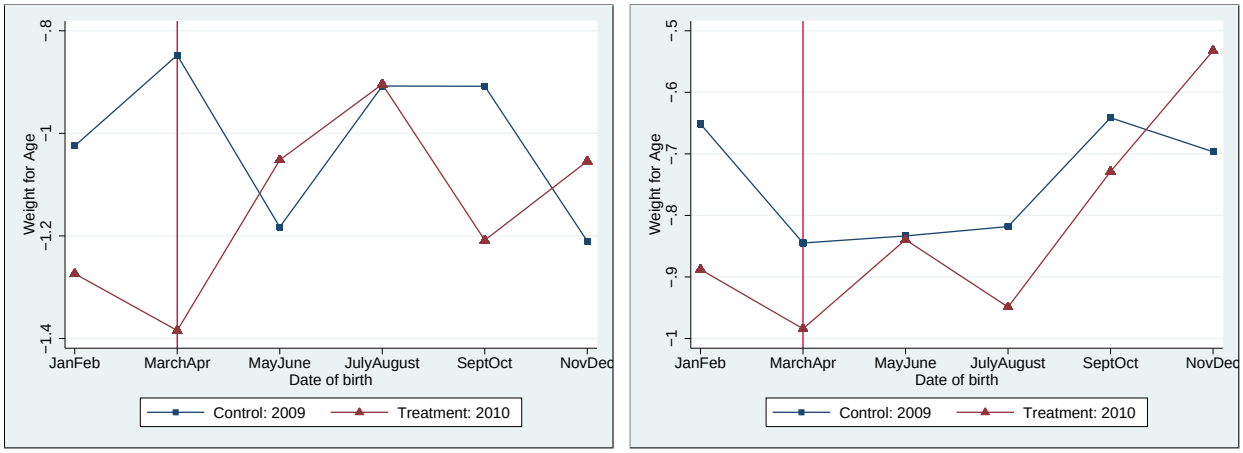
	Anthropometric Indicator		
	Treatment Group	*Control Group*	**Difference**
Panel A: Weight for Height			
Post-FHCI (May-December)	-0.147	-0.102	-0.045
	(0.059)	(0.066)	(0.081)
Pre-FHCI (January-April)	-0.181	0.011	-0.192**
	(0.070)	(0.081)	(0.103)
Difference	0.034	-0.113	**0.148**
	(0.086)	(0.131)	(0.125)
Panel B: Weight for Age			
Post-FHCI (May-December)	-0.909	-0.888	-0.021
	(0.068)	(0.071)	(0.087)
Pre-FHCI (January-April)	-1.117	-0.827	-0.290***
	(0.075)	(0.097)	(0.120)
Difference	0.207**	-0.061	**0.269***
	(0.086)	(0.147)	(0.140)
Panel C: Height for Age			
Post-FHCI (May-December)	-1.228	-1.324	0.096
	(0.079)	(0.083)	(0.115)
Pre-FHCI (January-April)	-1.598	-1.396	-0.202
	(0.097)	(0.101)	(0.139)
Difference	0.369***	0.072	**0.297**
	(0.126)	(0.131)	(0.181)

Notes: Each entry in cell reports the mean value of the anthropometric indicators in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Robust standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.



(a) Poor (b) Rich

Figure 3.7: Weight for Age: Poor vs Rich

seems to have increased the average weight for age of 0.45 standard deviations. The reason might lie in the initial disadvantageous position of such children, which allows larger room for improvement. The difference in results with respect to the choice of the place of delivery might be instead rooted into a higher propensity of poor mothers to take their child to hospital because of the absence of alternatives, while keeping on delivering at home assisted by the largely diffused traditional birth attendants. Overall, the results provide some evidence of a positive impact of the policy on poor children, which was considered as one of its primary aims.

Figure 3.8 and Table 3.9 report the effect separately for rural and urban children and show that it is positive but statistically not significant for neither group. This means that the overall effect is not guided by any of the two. Moreover, in terms of magnitude, the two groups do not differ considerably. Instead, as for the place of delivery, Figure 3.9 and Table 3.10 show how further children benefited more from the policy, probably for the reasons outlined in the previous subsection.

A new interesting aspect to investigate, in line with much literature concerned with assessing the impact of policies or programs in developing countries, is the so-called *gender effect*. It is widely observed that most advantages are enjoyed only among boys[24]. Only by comparing the trends for male and female children in Figure 3.10, it is clear how

[24]Tanaka (2014) shows how a similar policy in South Africa had higher beneficial effects for boys than for girls.

Table 3.8: Difference-in-differences Matrix: Weight for Age (Poor vs Rich)

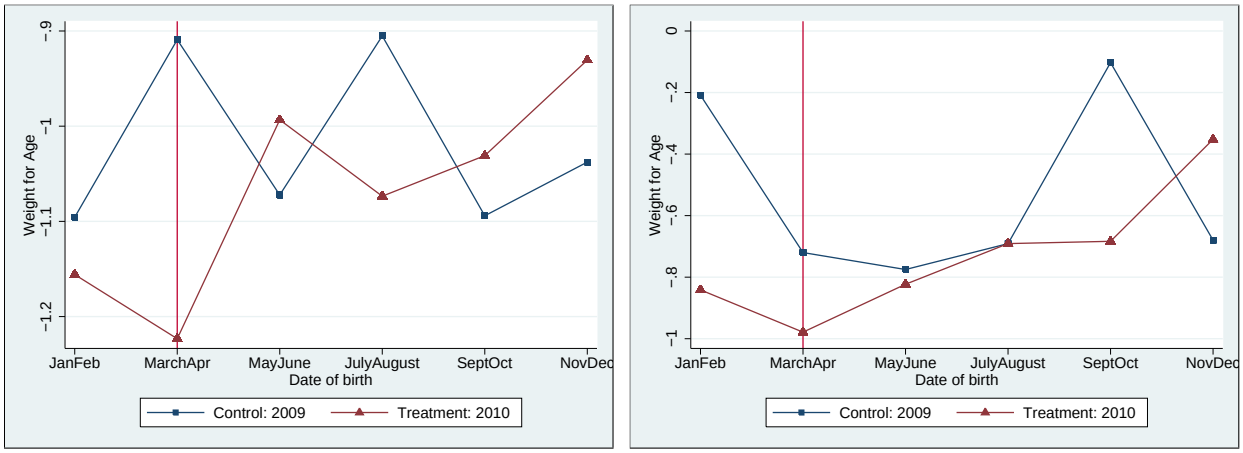
	Weight for Age		
	Treatment Group	*Control Group*	**Difference**
Panel A: Poor			
Post-FHCI (May-December)	-1.049	-1.063	0.014
	(0.086)	(0.095)	(0.128)
Pre-FHCI (January-April)	-1.341	-0.910	-0.431***
	(0.103)	(0.126)	(0.162)
Difference	0.292**	-0.153	**0.445****
	(0.136)	(0.155)	(0.207)
Panel B: Rich			
Post-FHCI (May-December)	-0.797	-0.763	-0.034
	(0.081)	(0.087)	(0.119)
Pre-FHCI (January-April)	-0.942	-0.762	-0.179
	(0.094)	(0.111)	(0.195)
Difference	0.145	-0.0008	**0.146**
	(0.126)	(0.140)	(0.188)

Notes: Each entry in cell reports the mean value of the anthropometric indicators in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Clustered standard errors are reported in the parentheses.

***Significant at the 1 percent level.

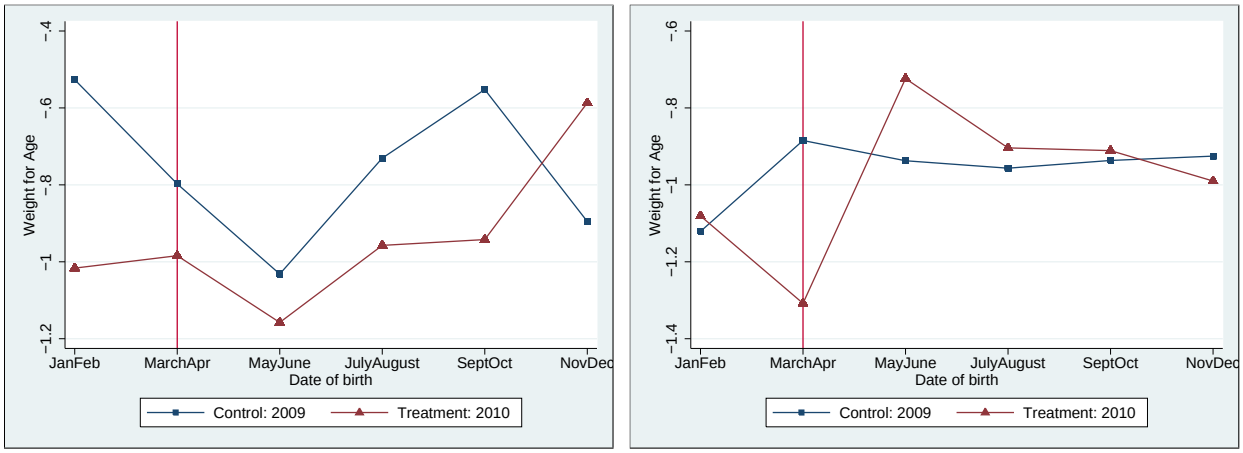
**Significant at the 5 percent level.

*Significant at the 10 percent level.



(a) Rural (b) Urban

Figure 3.8: Weight for Age: Rural vs Urban



(a) Close (b) Far

Figure 3.9: Weight for Age: Close vs Far

Table 3.9: Difference-in-differences Matrix: Weight for Age (Rural vs Urban)

	Weight for Age		
	Treatment Group	*Control Group*	**Difference**
Panel A: Rural			
Post-FHCI (May-December)	-1.010	-1.014	0.004
	(0.071)	(0.074)	(0.102)
Pre-FHCI (January-April)	-1.198	-0.983	-0.215*
	(0.77)	(0.092)	(0.120)
Difference	0.188*	-0.032	**0.219**
	(0.108)	(0.118)	(0.158)
Panel B: Urban			
Post-FHCI (May-December)	-0.682	-0.591	-0.091
	(0.104)	(0.127)	(0.164)
Pre-FHCI (January-April)	-0.906	-0.517	-0.389*
	(0.151)	(0.166)	(0.224)
Difference	0.224	-0.074	**0.298**
	(0.179)	(0.205)	(0.278)

Notes: Each entry in cell reports the mean value of the anthropometric indicators in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Clustered standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 3.10: Difference-in-differences Matrix: Weight for Age (Close vs Far)

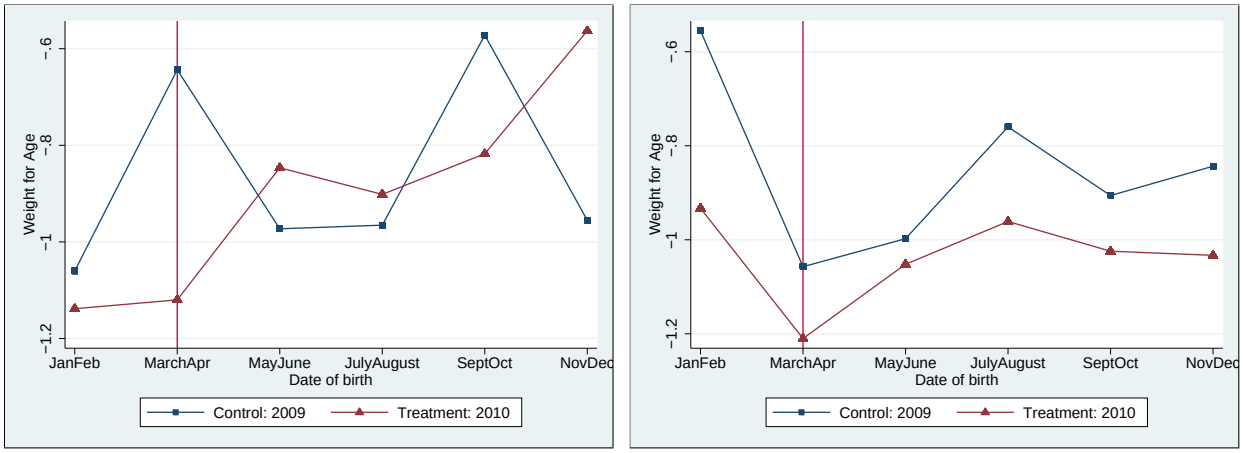
	Weight for Age		
	Treatment Group	*Control Group*	**Difference**
Panel A: Close			
Post-FHCI (May-December)	-0.953	-0.839	-0.114
	(0.077)	(0.087)	(0.117)
Pre-FHCI (January-April)	-0.999	-0.677	-0.322*
	(0.101)	(0.134)	(0.168)
Difference	0.046	-0.162	**0.208**
	(0.127)	(0.153)	(0.204)
Panel B: Far			
Post-FHCI (May-December)	-0.863	-0.942	0.079
	(0.090)	(0.096)	(0.132)
Pre-FHCI (January-April)	-1.223	-0.967	-0.256*
	(0.097)	(0.100)	(0.139)
Difference	0.360***	0.025	**0.335***
	(0.135)	(0.141)	(0.192)

Notes: Each entry in cell reports the mean value of the anthropometric indicators in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Clustered standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.



(a) Male (b) Female

Figure 3.10: Weight for Age: Male vs Female

the beneficial effect is male-oriented. The results in Table 3.11 reveal that the positive effects found in Panel B of Table 3.7 are mostly driven by boys: here, the point estimate is substantially larger than the previous estimates (0.416 versus 0.269) and highly significant. On the other hand, the magnitude of the effect, even if consistently positive, is lower for girls and not statistically different from zero. These differential effects cannot be explained by intrinsic biological gender differences, because health status looks similar between boys and girls, as can be directly observed in Figure 3.10. Instead, this is consistent with anecdotal evidence of son-preference in this kind of societies, and it indicates that even abolishing direct user fees will not necessarily benefit both genders equally.

Finally, Figure 3.11 and Table 3.12 report the results for children of low-educated mothers and of high-educated ones. The overall effect seems not to be driven by none of these two groups, for both of which the effect is statistically significant.

This subsection showed how the policy provided beneficial effects through the impact on children also for groups that seemed to be non-responsive in the choice of a health facility for delivering, namely poorer and less educated individuals. The following subsections will analyze the effect of the policy on mortality, both for children and for mothers.

Table 3.11: Difference-in-differences Matrix: Weight for Age (Male vs Female)

	Weight for Age		
	Treatment Group	*Control Group*	**Difference**
Panel A: Male			
Post-FHCI (May-December)	-0.804	-0.894	0.090
	(0.087)	(0.090)	(0.125)
Pre-FHCI (January-April)	-1.128	-0.802	-0.327**
	(0.104)	(0.119)	(0.157)
Difference	0.325**	-0.092	**0.416****
	(0.137)	(0.147)	(0.201)
Panel B: Female			
Post-FHCI (May-December)	-1.017	-0.883	-0.133
	(0.080)	(0.092)	(0.122)
Pre-FHCI (January-April)	-1.105	-0.852	-0.253*
	(0.095)	(0.117)	(0.151)
Difference	0.088	-0.031	**0.120**
	(0.125)	(0.148)	(0.194)

Notes: Each entry in cell reports the mean value of the anthropometric indicators in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Clustered standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 3.12: Difference-in-differences Matrix: Weight for Age (Low vs High Education)

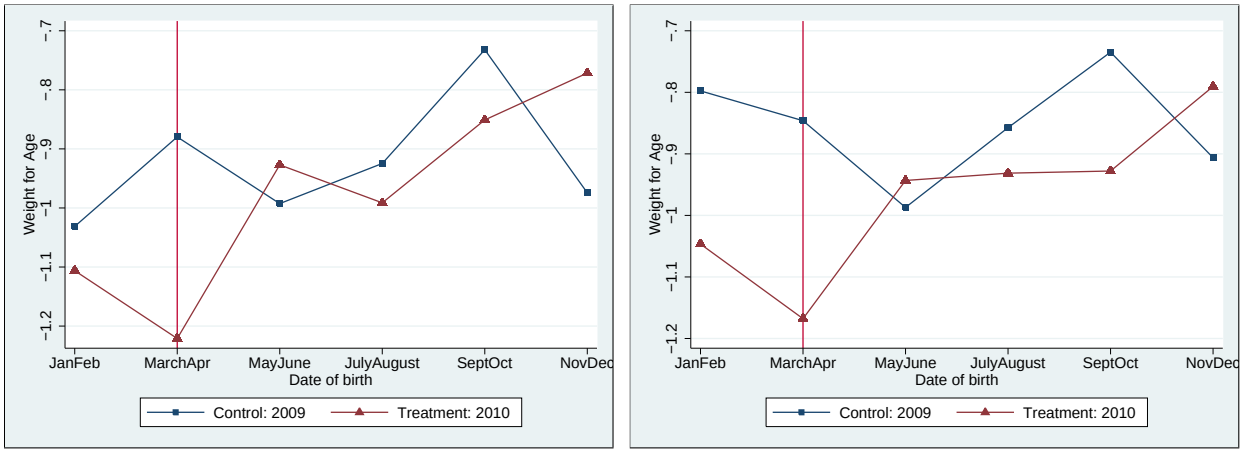
	Weight for Age		
	Treatment Group	*Control Group*	**Difference**
Panel A: Low Education			
Post-FHCI (May-December)	-0.900	-0.926	0.026
	(0.068)	(0.068)	(0.096)
Pre-FHCI (January-April)	-1.174	-0.943	-0.231**
	(0.071)	(0.084)	(0.110)
Difference	0.273***	0.017	**0.257***
	(0.101)	(0.108)	(0.146)
Panel B: High Education			
Post-FHCI (May-December)	-0.909	-0.888	-0.021
	(0.059)	(0.065)	(0.088)
Pre-FHCI (January-April)	-1.117	-0.827	-0.290 ***
	(0.070)	(0.083)	(0.109)
Difference	0.208**	-0.061	**0.269***
	(0.093)	(0.104)	(0.140)

Notes: Each entry in cell reports the mean value of the anthropometric indicators in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Clustered standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.



(a) Low Education (Mother) (b) High Education (Mother)

Figure 3.11: Weight for Age: Low vs High Education

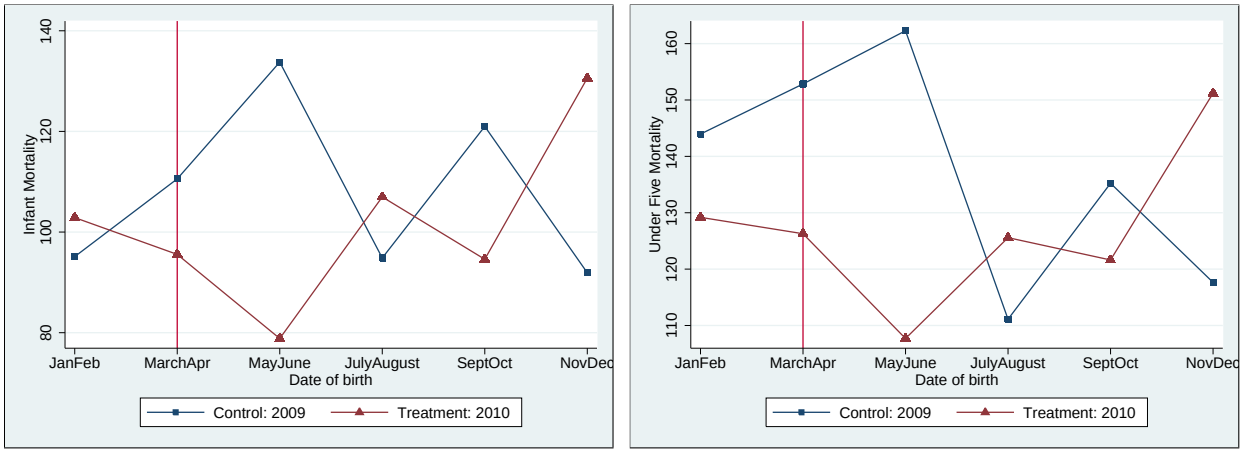
3.4.3 Infant and Under-Five Mortality

This section investigates the effect of the policy on child mortality. Two indexes are hereby proposed: infant mortality and under five mortality. They are usually expressed in terms of number of deaths every 1,000 live births as follows:

$$INFANT = \frac{number\ of\ deaths\ within\ 12\ months\ of\ age}{number\ of\ live\ births} * 1000 \tag{3.4}$$

$$UNDERFIVE = \frac{number\ of\ deaths\ within\ 5\ years\ of\ age}{number\ of\ live\ births} * 1000 \tag{3.5}$$

Only by observing the graphs in Figure 3.12, the lack of a clear effect of the policy on the outcomes of interest seems quite straightforward. Indeed, they do not show a desirable decreasing trend after April none of the outcomes: right after June the ratio starts to increase again. The policy seems to not have had any short-run effect on mortality for children. Again, as underlined elsewhere, no conclusion can be drawn for the long run, for which the identification of the causal effect would become problematic.



(a) Infant (b) Under Five

Figure 3.12: Mortality (Infant and Under Five)

3.4.4 Maternal Mortality

One of the primary aims of the FHCI was to decrease maternal mortality, whose level is in Sierra Leone among the highest in the world.

There exist several indexes that are capable to summarize information relative to maternal mortality. One of these, which is used in this analysis, is the so called maternal mortality ratio (MMRatio) which is defined as the ratio between the number of maternal deaths within a specified period and the number of live births within the same period. It is usually expressed per 100,000 live births[25]:

$$MMR = \frac{number\ of\ maternal\ deaths}{number\ of\ live\ births} * 100.000 \tag{3.6}$$

As for the numerator, the WHO defines maternal deaths as follows: *A maternal death is the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and the site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental cause*[26].

[25]The main reference for how to exploit the DHS dataset for maternal mortality is Stanton et al. (2000). For an overview of the sisterhood method for estimating the maternal mortality ratio, see Graham et al. (1989).

[26]WHO, 2004.

In order to asses the effect of the program on maternal mortality it is necessary to estimate the MMR for each month for the identification of the treatment effect. The DHS dataset embodies information necessary for an estimation of maternal mortality: in the individual woman's questionnaire, respondents are asked to list their siblings and the respective age and sex, thus reconstructing their mothers' birth history. Moreover, sex and date of death of the deceased sibling are provided together with information regarding the cause of death for the female sibling, in particular the existence of some link to pregnancy, childbirth or the postpartum period.

The method employed for estimating the MMR from such kind of information is called the *sisterhood method*. It is a relatively new technique, aiming at deriving population-based estimates of maternal mortality using the proportions of adult sisters dying for reasons linked to pregnancy or delivery reported by adults during a census or survey like the DHS (Graham et al. 1989). In order to compute the short-run impact of the FHCI on maternal mortality, I estimate a MMR for every month as follows. The numerator of Formula 3.6, i.e. the number of maternal deaths, is computed by counting the number of women that died either *while pregnant*, *during delivery* or *two months after delivery*, as expressed in the individual recode (whose unit of observation is the adult single individual). Instead, the birth recode is used as base for an estimation of the denominator, i.e. for the computation of the number of live births for each month considered. Hence, I compute the ratio by dividing the number of maternal deaths in every month for the live births of that specific period, merging the birth recode with the individual recode. In order to test the robustness of such estimate, I compare the average Maternal Mortality Ratio over the period 2007-2010 computed through the process above to official estimates. I compare the average MMR for the period 2007-2010 (MMR=1741) to the estimate provided by the DHS final report for the period 2006-2013 (MMR=1165) or the World Bank estimate for 2013 (MMR=1200)[27].

As for the other indicators so far analyzed, the average MMR for each group of months for the years 2007, 2008, 2009 constitute the control group[28], whereas the MMRs

[27]The latter ratio is estimated with a regression model exploiting information on the proportion of maternal deaths among non-AIDS deaths in women ages 15-49, fertility, birth attendants, and GDP. The average MMR in this study is higher than both the ratios reported. This discrepancy might be due to a difference in the periods covered, other than due to the difference in the methodology adopted.

[28]For data availability reasons, the control group considers only cohort of women died during three years before the policy, instead of six.

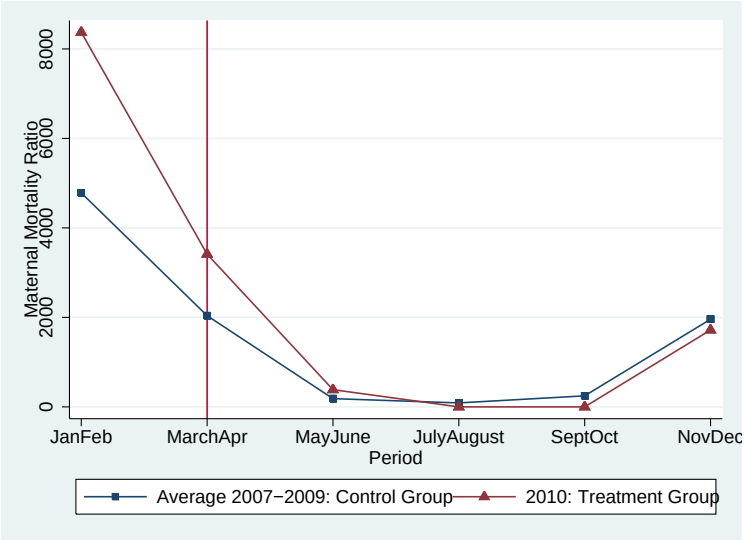


Figure 3.13: Maternal Mortality Ratio: Trends

for 2010 represents the treatment group. Figure 3.13 show trends for both groups, and Table 3.13 reports the difference-in-differences matrix with the treatment effect.

The table suggests a decrease in the MMR of around 2,600 deaths every 100,000 live births, a desirable effect. Nevertheless, the coefficient is not statistically significant. Therefore, a short-run effect of the policy on maternal mortality is not statistically confirmed and remains ambiguous. However, a positive effect is here more plausible than in the case of infant and under-five mortality, as evident from the graphs in Figure 3.13 and from the fact that within the treatment group, the difference between the post-FHCI and the pre-FHCI average MMR is negative and statistically significant[29].

Despite a non-significance of the overall effect, I repeat the comparison for poor/rich, rural/urban and close/far mothers, whose trends are reported in Figure 3.15 and Figure 3.14 (results for close/far mothers are not shown, and the reason is mentioned below). All the results are statistically non-significant, but the magnitudes of the beneficial effects seem to be higher for poor, urban women and mothers close to a hospital, results that partially support the considerations made so far. Only the fact that close mothers seem to have benefited more is somehow new. However, the result should be considered carefully. Apart from the already underlined issues concerning the problematic aspects

[29]So it is for the control group, and this might imply that the difference is only due to seasonality effects.

Table 3.13: Difference-in-differences Matrix: Maternal Mortality Ratio

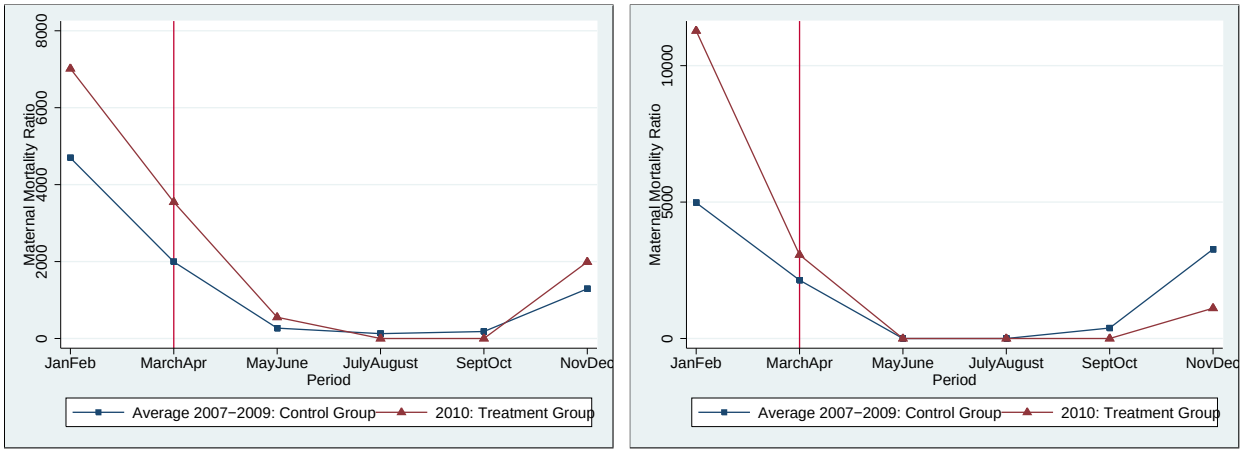
	Maternal Mortality Ratio		
	Treatment Group	*Control Group*	**Difference**
Post-FHCI (May-December)	500	597	-97
	(404)	(451)	(606)
Pre-FHCI (May-June)	5927	3420	2507
	(1895.0)	(984.9)	(2135.7)
Difference	-5427***	-2823**	**-2604**
	(1456.7)	(945.2)	(2220.0)

Notes: Each entry in cell reports the mean value of the maternal mortality ratio (computed for each single month in the year) for the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Robust standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.



(a) Rural (b) Urban

Figure 3.14: Maternal Mortality Ratio : Rural vs Urban

linked to distance, another element here has to be taken into account. Indeed, the DHS survey does not provide direct information concerning the specific mothers that have died, but only about the respondent, i.e. the sibling, and these pieces of information are the ones used for these comparisons. If one can assume that some characteristics of the sibling might have been shared by the sister (income, for example, or maybe also the categorization as rural or urban, even though this assumption is already quite strong), some other dimensions are less likely to be the same, among which the place of residence (which makes the mother fall into the category close/far from a hospital).

To sum up, the short run effect on maternal mortality is ambiguous, even though slightly more pronounced than for infant and under-five mortality. One can speculate and believe that this is due to a mother-biased practice in hospitals (which means that mothers are considered first, when both in danger), but the non-significance of the coefficient for maternal mortality leaves the issue unsolved[30].

3.5 Robustness Checks

The aim of this section is to provide some evidence for the robustness of the results related to the choice of the place of delivery and to the weight for age z-score by relaxing the

[30]Other studies have provided ambiguity in the effect of such policies on mortality. See Bosu et al. (2007).

Table 3.14: Difference-in-differences Matrix: Maternal Mortality Ratio (Poor vs Rich)

	Maternal Mortality Ratio		
	Treatment Group	*Control Group*	**Difference**
Panel A: Poor			
Post-FHCI (May-December)	382	506	-124
	(200)	(400)	(447)
Pre-FHCI (January-April)	6188	3154	3034
	(1756)	(748)	(1909)
Difference	-5806***	-2648***	**-3158**
	(1273)	(774)	(1960)
Panel B: Rich			
Post-FHCI (May-December)	617	665	-48
	(632)	(503)	(808)
Pre-FHCI (January-April)	5993	3625	2368
	(2783)	(1276)	(3062)
Difference	-5376**	-2961**	**-2416**
	(2162)	(1149)	(3167)

Notes: Each entry in cell reports the mean value of the maternal mortality ratio (computed for each single month in the year in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Robust standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 3.15: Difference-in-differences Matrix: Maternal Mortality Ratio (Rural vs Urban)

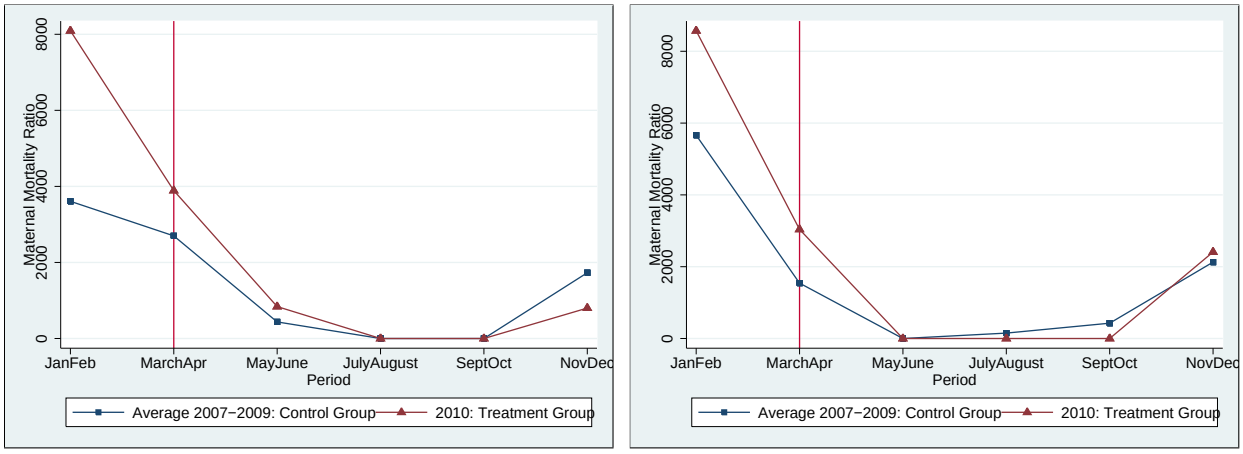
	Maternal Mortality Ratio		
	Treatment Group	*Control Group*	**Difference**
Panel A: Rural			
Post-FHCI (May-December)	590	449	140
	(450)	(288)	(534)
Pre-FHCI (January-April)	5391	3333	2058
	(1477)	(1028)	(1799)
Difference	4801***	2883***	**-1918**
	(1223)	(834)	(1877)
Panel B: Urban			
Post-FHCI (May-December)	284	889	-605
	(291)	(792)	(844)
Pre-FHCI (January-April)	7273	3633	3640
	(4278)	(1082)	(4412)
Difference	6989**	2744*	**-4245**
	(3052)	(1356)	(4492)

Notes: Each entry in cell reports the mean value of the maternal mortality ratio (computed for each single month in the year in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Robust standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.



(a) Poor (b) Rich

Figure 3.15: Maternal Mortality Ratio: Rich vs Poor

functional form assumption and estimating the treatment effect non-parametrically[31].

One of the main problems in any microeconometric evaluation study of any kind is to address the so-called selection bias. Indeed, when comparing a treatment and a control group, what can occur is that the control group does not perform well in simulating the behavior of the treatment group, because of differences in observed and not-observed characteristics[32]. The matching approach is one of the possible ways to address the selection problem. The basic idea is to find in a large group of non-participants[33] those individuals that are similar to the participants in all relevant pre-treatment characteristics X. Once this is done, differences in outcomes between this well-selected and adequate control group and the treatment group can be attributed to the policy itself.

A parsimonious way of implementing this procedure is making use of the propensity score, which is the probability of participating in a program given observed characteristics X. This probability is usually estimated through a probit model, that has as dependent variable the dummy treated/untreated and as explanatory variables the just mentioned

[31] See Chapter 1.

[32] Since I have adopted cohorts of mothers that have delivered (or children born) in years preceding the initiative, one can claim that treatment and control group are likely to be similar in terms of observables and unobservables, since mothers do not decide when to deliver or children when to get born, i.e. in which group to fall, whether treated or non-treated. Treatment and control groups are as if they were randomized. This section will indeed prove that results are not driven by observable characteristics distinguishing the two groups, but only by the change of policy.

[33] The advantage of this context is that there are many observations available.

characteristics X. After this, observations are matched according to their propensity score through a matching algorithm[34].

I choose as covariates for the computation of the propensity score the observed characteristics so far adopted, namely the wealth index, the place of residence (rural or urban), the years of education and distance to a health facility[35]. The matching algorithm I adopt is a kernel matching, a non-parametric matching estimator that uses weighted averages of all the individuals in the control group to construct the counterfactual outcome[36].

In the non-parametric world, one of the first elements that have to be assessed is the so called *common support assumption*. In simple words, it ensures that individuals with the same X values have a positive probability of being both participants and non-participants. Let P be the propensity score, or the probability of falling into the treatment group, X the vector of covariates and *Treated* the dummy that describes treatment. The common support assumption requires that:

$$0 < P(Treated = 1|X) < 1 \tag{3.7}$$

The first step I implement for assessing the validity of such assumption is computing the propensity score through a probit regression as described above, and then plotting the density of the propensity score for treated and not treated, in order to visually identify the regions of common support. After that, I compute the treatment effect through a kernel-weighted propensity score matching difference-in-differences strategy, restricting to common support. The following subsections report separately the robustness checks for the effect on the place of delivery, on the one hand, and for the weight for age z-score, on the other.

[34]For a detailed list of such algorithms, see Caliendo & Kopeinig (2008).

[35]In the choice of variables to include in the computation of the propensity score, Caliendo & Kopeinig (2008) recommend to consider only variables that are unaffected by participation. One can argue that both distance to a health facility and the place of residence cannot be excluded by the set of variables that the policy might have influenced. However, since we are in a short-run framework, such characteristics are not likely to change so quickly.

[36]Weights are associated according to a symmetric, nonnegative and unimodal kernel, which tends to attach higher weights to individuals in the control group that are closer in terms of propensity score to individuals in the treatment group. See Chapter 1.

3.5.1 Delivering in a Health Facility

Figure 3.16 compares the densities of the propensity score of the control group (on the left) to the one of the treatment group (on the right). The distributions look quite similar, and this confirms that the two groups are, as conjectured so far, similar in terms of the observable characteristics considered. However, there are some regions for which the common support assumption does not hold, where only observations for the control group are available: between 0.3 and 0.35. For this reason, I estimate the treatment effect by a kernel propensity score matching difference-in-differences strategy restricting to common support, i.e. eliminating observations that fall outside the overlapping region, and report the result in Table 3.16. The bandwidth I adopt is 0.06, but the results are robust to the choice of bandwidth (not reported). Moreover, standard errors are bootstrapped, thus allowing proper inference, that could not be applied with traditional standard errors because they would not take into account the estimation of the propensity score. As can be noticed, the effect is almost identical to the one identified in Section 3.4.1, which is thus robust to alternative specifications, in particular to the relaxation of the assumption concerning the functional form, typical of parametric models as the one adopted in this study.

3.5.2 Weight for Age z-score

The comparison of the density of the p-score for the treatment and control group is reported in Figure 3.17. The control group here is different from the one analyzed in the previous subsection and the variable *Gender* is included in the estimation of the propensity score, being, as seen, a relevant characteristic for the distribution of the effect. The densities look quite similar, despite some tiny areas where the common support assumption fails. Again, I implement the same procedure as in the previous section and report the results in Table 3.17. As can be seen, the latter suggests a positive and significant effect of 0.27 standard deviations. This proves that the results proposed in Section 3.4.2 are robust to a relaxation of functional form assumptions.

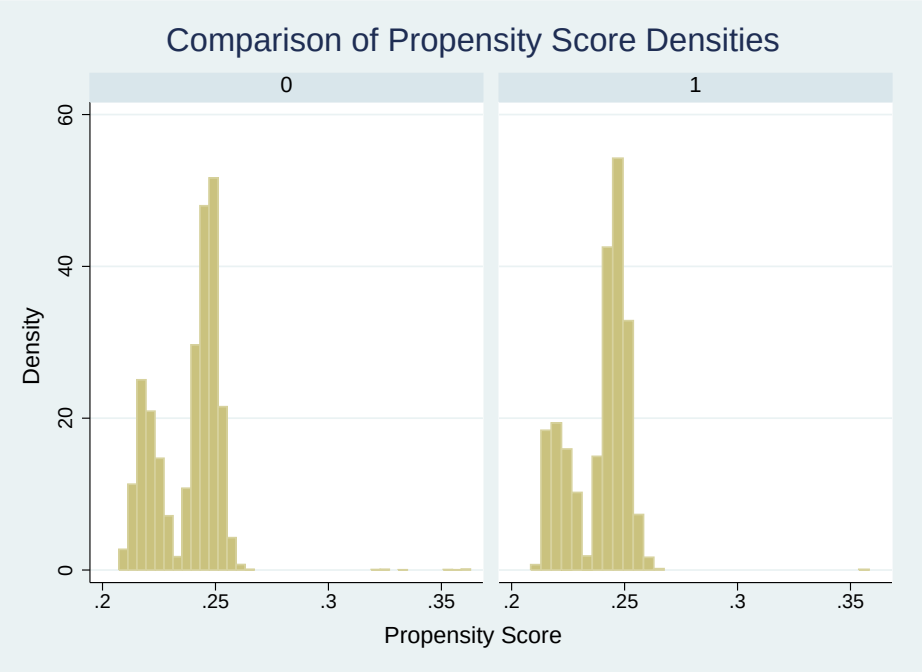


Figure 3.16: Propensity Score Density: Control Group (left), Treatment Group (right).

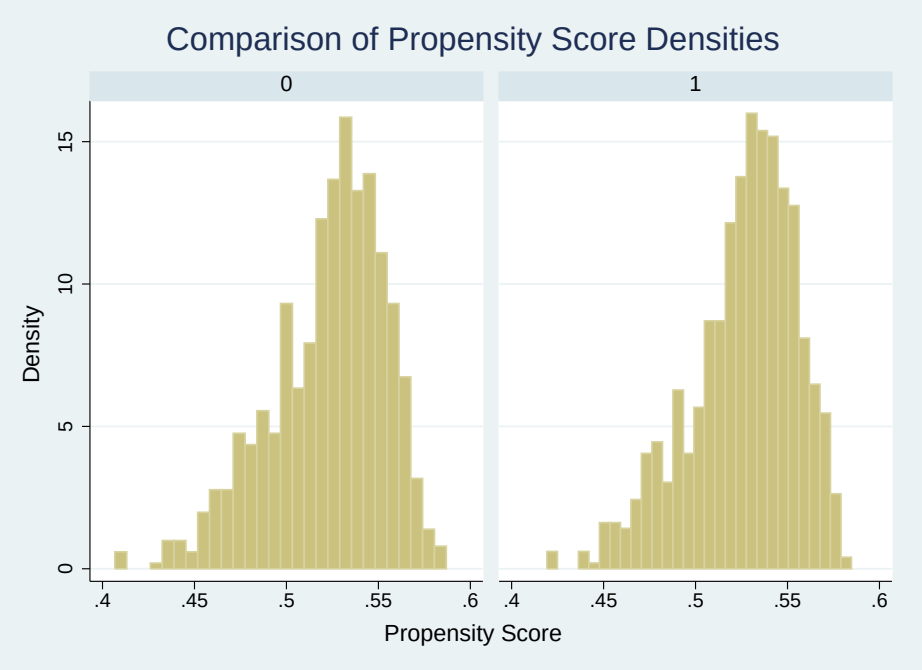


Figure 3.17: Propensity Score Density (Anthropometry): Control Group (left), Treatment Group (right).

Table 3.16: Difference-in-differences Matrix: Delivering in a Governmental Health Facility (Propensity Score Matching)

	Frequency of women delivering in a health facility		
	Treatment Group	*Control Group*	**Difference**
Post-FHCI (July-December)	0.521	0.314	0.206***
	(0.018)	(0.007)	(0.021)
Pre-FHCI (January-June)	0.445	0.282	0.163***
	(0.010)	(0.007)	(0.012)
Difference-in-differences			**0.043***
			(0.023)

Notes: Bootstrapped standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 3.17: Difference-in-differences Matrix: Weight for Age (Propensity Score Matching)

	Weight for Age		
	Treatment Group	*Control Group*	**Difference**
Post-FHCI (July-December)	-0.909	-0.896	-0.013
	(0.057)	(0.063)	(0.080)
Pre-FHCI (January-June)	-1.124	-0.847	-0.278**
	(0.080)	(0.087)	(0.134)
Difference-in-differences			**0.264***
			(0.137)

Notes: Bootstrapped standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Dicussion and Conclusions

This study has tried to address the topic of health user fees by comprehensively review the main points of the complex debate on it and proposing a contribution through a case study for a specific country, Sierra Leone. To this end, the most up-to-date empirical methods have been adopted, adapting them to the available data and the context under study.

The Free Health Care Initiative introduced in Sierra Leone in 2010 has had a positive impact on the demand for health care for the target categories, effect identified by many other studies investigating the impact of similar policies in other contexts, as largely illustrated. Thanks to the elimination of health user fees, the probability for a woman to deliver in a governmental health care facility has been shown to have increased by 4.4 percentage points in the months immediately after the policy, thus in a short-time horizon. As in many other contexts, positive effects have been recorded for an overall indicator of nutrition and health for children under five, such as weight for age. As underlined, a desirable effect on anthropometric indicators suggests the presence of both a direct positive effect, especially in the first months of life, and an indirect effect, through the benefits received by the mothers, exposed to the program during pregnancy and the lactating period, when they are partially if not totally responsible not only for the decision of taking children to hospital if sick, but also for their nutrition through breastfeeding.

If some benefits are clearly visible, suggesting that the policy has been effective in reaching some goals, at least in the short run, this study highlights some important challenges. First of all, ambiguous are the effects of the initiative on child mortality, and partially, on maternal mortality, when one of the crucial aims was reducing exactly these indicators, as stated in the Millennium Development Goals of 2015. As seen, the study focuses on the short run, and one can claim that the responsiveness of such indicators is slower than the one of demand and anthropometry. Therefore, a possible long-run

effect of the policy on mortality cannot be excluded on the basis of the results presented, even though it would be hard to prove such effect to be related to the Free Health Care Initiative. Nevertheless, in investigating the different responsiveness of some categories, this study sheds light on some elements that deserve to be discussed.

First of all, as shown for the frequency of women delivering in a health care facility, the effect on lower-income groups seems not to have been as high as one would have expected, in light of some evidence concerning the higher elasticity of health-care demand for poorer households, presented in Chapter 2. Some interpretation for this result is provided in Section 2.2 and in the Appendix and will not be further stressed. However, what one can infer is that removing financial barriers is not a sufficient measure for fostering demand for health-care in low-income groups. The straightforward answer to this puzzling issue might be claiming that additional indirect costs could be the source of such a low responsiveness.

Nevertheless, this study provides evidence of higher effects for mothers and children that reside far away from health facilities, implying that an indirect cost such distance becomes less relevant once economic barriers are removed. Therefore, what should be put under the attention of policy makers is the cultural and educational barriers in access to health care: not only the effect is lower for less educated mothers, but once controlling for education the difference in the effect between rich and poor is not statistically significant anymore. If the goal of targeting sensitive categories such as mothers and children can be said to have been partially reached, targeting the lowest quintiles of the population seems to be a harder task. Additional measures such as educational policies aiming at increasing awareness about the importance of health care among the poorest are needed. The evidence provided by this study supports the claim of a whole stream of literature on developing countries insisting on conceiving education and health as elements mutually reinforcing each other. In this particular context, the attention should be given to implementing measures for further increase the ratio of women receiving skilled care during delivery. Particularly in rural areas, the phenomenon of delivering at home assisted by a traditional birth attendant is still widespread, as confirmed by the higher impact of the policy on urban population. This is only one manifestation of the tendency of turning to traditional healers whenever an illness occurs. The recent Ebola outbreak and the channels through which the virus has spread just confirm these claims: instead of referring

to health centers, many infected individuals turned to traditional healers[37], because of misconceptions and fear (typical of animist tribes like the ones in Sierra Leone), allowing the virus to diffuse more quickly. However, the responsiveness of poor children in terms of weight-for-age is encouraging, leading to conclude that for some kind of health services removing financial barriers might turn out to be beneficial also for low-income households.

An important and further caveat has to be highlighted, when reasoning about the removal of health user fees in developing countries. As stressed in Chapter 2, one of the immediate shortcoming of such kind of policies is long term sustainability. Much aid has been devolved for the implementation of the Free Health Care Initiative, given that the natural consequence of the removal of user fees is the reduction in revenues for the public health sector. In a country where public spending is particularly low, in the long run this revenue-loss might be the source of critical inefficiencies in guaranteeing proper health services. One way of addressing this issue might be differentiating fees not only across categories, as in the Free Health Care Initiative, but also across income groups. In order to do this, however, two challenges would arise: identifying the income group to which patients belong and avoiding a shift towards the private sector, on average providing higher quality service. Moreover, financial support from international partners is likely to not be persisting in the long run. All these factors are likely to be threatening the quality of service, particularly under pressure during the last year because of the Ebola virus outbreak.

Finally, some weaknesses of this study have to be underlined. The lack of facility-based data has not allowed to investigate important aspects like the impact of such a policy on quality of health-care facilities, which has also been proven to be a relevant factor in the choice of health provider in a considerably wide stream of literature. Moreover, the availability of such type of data would have been useful in order to assess in a more detailed manner the impact on demand, analyzing the effect on services such as vaccinations, antenatal or postnatal care. This can only be indirectly inferred, with the kind of data used in this study, where only outcomes that can be reconciled to a temporal dimension can be exploited. Moreover, a study of this kind is likely not to have strong external validity. The sample here analyzed can be considered to be representative for Sierra Leone as a whole, as underlined when reporting the sampling techniques, but the

[37]One of the reasons of the institution of curfews all over the country was exactly the movement of people towards the more well-known traditional healers.

conclusions drawn in this study can not be extended beyond the boarders of the country under consideration. Additionally, as more than once underlined, no inference can be drawn with respect to a long run effect of the policy from the results here discussed. This aspect can also be seen as a limitation, since a change as the FHCI is likely to produce effects only in the long run, as far as indicators like mortality are concerned.

To conclude, changes in financing policy or in other aspects of health system design that are implemented countrywide are hard to be evaluated with a rigorous procedure like in a randomized experiment. Many factors are likely to change at the same time and trying to establish causality is often a challenging task. If the strategy adopted in this study highlighted positive effects for demand and health outcomes for children, many issues remain instead unsolved, such as identifying the long run effects of such a policy. The open questions that should be further investigated relate to finding possible ways to target more directly the poor and to assess the long-term sustainability of a policy like the Free Health Care Initiative.

Appendix

Delivering in a Governmental Health Facility: Effect on Poor and Rich Women

As discussed in Section 3.4.1, the FHCI seems to have had higher effects on richer mothers, a result that seemed, at a first sight, counterintuitive. In this respect, the hypothesis that one can formulate is the following: richer women are likely to be more educated and living in urban areas, thus belonging to categories that revealed themselves to be more responsive to the policy change (see again Sections 3.4.1).

This hypothesis concerning the characteristics of poor and richer mothers can easily be tested through simple summary statistics. Table 4.1 reports the frequency of women living in urban areas separately for richer and poorer women, together with the mean value of their respective years of education. Roughly 50% of richer women live in urban areas. The percentage is significantly different from the one for the poorer women in the sample, of which 95% are instead living in rural areas. Additionally, the two groups differ significantly with respect to years of education. On average, richer mothers exhibit almost two additional years of education compared to poor women. The Table reports a t-test for assessing the significance of the difference across groups. For both indicators, the difference is statistically significant.

Another viewpoint from which to analyze the issue is provided by Table 3.19, in which the coefficients of correlation between income quintile and place of residence, on the one hand, and between income quintile and years of education, on the other are reported. Both coefficients are positive, and this suggests that the higher the income quintile (the richer the mother) the highest the probability of residing in a urban area and the highest the years of education.

If this hypothesis is correct, it is left to show that once controlling for these two covariates, namely place of residence and years of education, the difference in the effect for poorer and rich women either goes on the other directions or is not statistically significant anymore. In order to test this, I run the following regression:

$$Y_i = \beta_0 + \beta_1 Treat_i + \beta_2 After_i + \beta_3 Treat_i \cdot After_i + \beta_4 Treat_i \cdot After_i \cdot Poor_i + \beta_4 X_i + \epsilon_i$$

where *Poor* is a dummy variable that takes value 1 if individual i belongs to the lowest 2 income quintiles and X is the vector of controls (years of education and place of residence). The coefficient β_4 captures the difference in the effect between rich (for which the effect is shown by β_3) and poor mothers.

Table 3.20 shows the regression results. As can be seen, once education and place of residence are accounted for, the effect on poor is not statistically significant anymore, even though the difference (β_4) persists on being negative. This supports the hypothesis of Section 3.4.1. What has to be underlined, though, is that the conclusion for which poor have not benefited more from the abolition of health user fees still holds. The FHCI has failed in targeting the poor categories of the population.

Anthropometric Indicators

One of the aim of this study was to verify whether the FHCI had some impact not only on demand for health care, but also on health outcomes. A useful tool in this respect, widely adopted in studies of this kind, is the use of anthropometry, because inexpensive and non-invasive measure of the general nutritional status of an individual or a population group (Cogill 2003). Childhood under-nutrition results from the synergistic effects of repeated and improperly treated illnesses and inadequate food intake. The weight-for-age, height-for-age and weight-for-height indexes are the most used for summarizing the nutritional and anthropometric status of a child. As underlined in Section 3.4, this variable is particularly suitable for the purpose of the study, since it allows to incorporate the double effect of the FHCI on health and nutrition simultaneously, assuming that a healthier mother is more likely to breastfeed the child.

As seen, these indicators are usually expressed in terms of z-scores or standard deviation scores from the average (or median) of the International Reference Population.

Table 3.18: Summary Statistics for Poorer and Richer Mothers

	Poorer Mothers	*Richer Mothers*	**Difference**
Ratio of mothers living in a urban area	0.05 [0.49]	0.51 [0.21]	-0.46*** (0.006)
Average years of education	1.1 [3.7]	3.1 [5.5]	-1.9*** (0.07)
Number of observations	7284	9963	

Notes: Each entry in cell reports the frequency of women living in a urban area or the average years of education in the corresponding groups (richer vs. poorer). Standard deviations are reported in squared brackets, standard errors in the parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 3.19: Correlation Coefficients: Income - Place of Residence;
Income - Years of Education

	Dummy for Urban	*Years of education*
Income quintile	0.65	0.33
Number of observations	17247	17247

Notes: Each entry in cell reports the correlation coefficient between the income quintile to which the individual belongs and the place of residence (expressed by a dummy that takes value of 1 if urban) and years of education.

Table 3.20: Regression Results: Controlling for Place of Residence and Education

	Frequency of women delivering in a governmental health facility
Treat	0.165***
	(0.0184)
After	0.0271**
	(0.0109)
Treat·After	0.0509*
	(0.0268)
Poor·Treat·After	-0.0107
	(0.0354)
Rural	-0.160***
	(0.0213)
Years of Education	0.0115***
	(0.00176)
Cnstant	0.370***
	(0.0187)
N	10492
adj. R^2	0.071

*Notes:*The table reports the regression results for the estimation of the treatment effect simultaneously for poorer and richer mothers. The coefficients of interest are the ones for Treat · After and Poor · Treat · After. Clustered standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

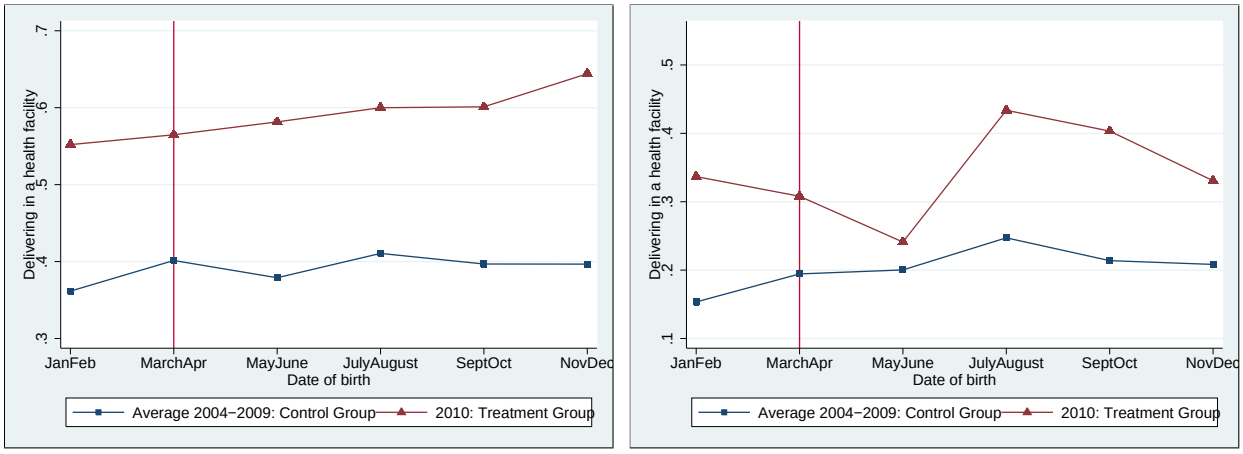
The reference standards most commonly used to standardize measurements were developed by the US National Center of Health and Statistics (NCHS) and are recommended for international use by the WHO. The reference population chosen by NCHS was a statistically valid random population of healthy infants and children. One can wonder whether the US-based NCHS reference standards is valid for populations from other ethnic backgrounds. Evidence suggests that until the age of approximately 10 years, children from well-nourished and healthy families throughout the world grow at approximately the same rate and attain the same height and weight as children from industrialized countries (Cogill 2003). References are used to standardize a child's measurement by comparing the child's measurement with the median or average measure for children at the same age and sex (in the DHS dataset, the median). This allows direct comparisons across groups and especially across cohorts of children with different age and sex (as in this study), and any desirable effect is identifiable with an increase in the indicator (the unity of measure is in the DHS the standard deviation). In short, the z-score is thus computed:

$$Z - score = \frac{\text{(observed value) - (median reference value)}}{\text{standard deviation of reference population}} \tag{3.8}$$

The DHS dataset used in this study is particularly precise in the methodology of anthropometric data collection. Weight and height for children between 0 and 5 years of age are collected as recommended by the WHO (these indicators are not suitable for considerations of health and nutrition for older children).

Results with a Self-reported Measure of Distance

This section replicates the results on the effect of the FHCI on demand exploiting a self-reported measure as an indicator of distance from a health facility, expressed by the respondent itself when asked about the role of distance when it comes to seek care in a health facility (whether it constitutes a big problem, not a big problem or no problem at all). This is needed in order to check for the consistency of the results obtained with the geodetic distance in Section 3.4. Figure 3.18 shows the trends for close and far groups in the frequency of women delivering in a governmental health facility. The patterns seem to mirror what encountered in Section 3.4. Farther women look as to have been more responsive: the jump after June is considerable, whereas no notable increase can be



(a) Close (Self-reported) (b) Far (Self-reported)

Figure 3.18: Delivering in a Governmental Health Facility: Self-reported Measure of Distance

detected for closer women (or, in other words, for those who perceive distance to be not a problem or a big problem when seeking care). Additionally, what is here particularly interesting to notice is the overall lower rate of far women delivering in hospital compared to the closer ones. This feature in the trends was not recorded in Section 3.4 with geodetic distances. Instead, what observed here is more consistent with the plausible assumption that women are in general less likely to seek care in a hospital, if living far from it. The FHCI, abolishing part of the economic direct cost, looks as to have encouraged women living farther away to deliver in a health facility, as already underlined. Table 3.21, reporting the difference-in-differences matrix, confirms what said so far and mirrors the effects encountered in Section 3.4.

Table 3.21: Difference-in-differences Matrix: Delivering in a Governmental Health Facility

	Frequency of women delivering in a health facility		
	Treatment Group	*Control Group*	**Difference**
Panel A: Close (Self-Reported)			
Post-FHCI (July-December)	0.613	0.401	0.212***
	(0.025)	(0.017)	(0.027)
Pre-FHCI (January-June)	0.567	0.383	0.184***
	(0.024)	(0.015)	(0.025)
Difference	0.045*	0.018	**0.028**
	(0.026)	(0.015)	(0.030)
Panel B: Far (Self-reported)			
Post-FHCI (July-December)	0.393	0.224	0.170***
	(0.029)	(0.014)	(0.029)
Pre-FHCI (January-June)	0.294	0.182	0.111***
	(0.025)	(0.012)	(0.025)
Difference	0.099***	0.041***	**0.058***
	(0.029)	(0.013)	(0.033)

Notes: Each entry in cell reports the mean value of the frequency of women delivering in a health facility (computed for each single month in the year as the number of women going to hospital over the total of mothers delivering) in the corresponding groups (treatment vs. control) and periods (pre- vs. post-reform). Clustered standard errors are reported in the parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Bibliography

Akin, J. S., Griffin, C. C., Guilkey, D. K. & Popkin, B. M. (1986), 'The demand for primary health care services in the Bicol region of the Philippines', *Economic Development and Cultural Change* pp. 755–782.

Akin, J. S., Guilkey, D. K., Hazel, E. et al. (1995), 'Quality of services and demand for health care in Nigeria: A multinomial probit estimation', *Social Science & Medicine* **40**(11), 1527–1537.

Angrist, J. D. & Krueger, A. B. (1999), 'Empirical strategies in labor economics', *Handbook of Labor Economics* **3**, 1277–1366.

Arthur, E. (2012), 'Wealth and antenatal care use: Implications for maternal health care utilisation in Ghana', *Health Economics Review* **2**(1), 1–8.

Asante, F. A., Chikwama, C., Daniels, A. & Armar-Klemesu, M. (2007), 'Evaluating the economic outcomes of the policy of fee exemption for maternal delivery care in Ghana', *Ghana Medical Journal* **41**(3).

Bank, W. (1987), *Financing Health Services in Developing Countries: An Agenda for Reform*, Washington DC.

Barber, S., Bonnet, F. & Bekedam, H. (2004), 'Formalizing under-the-table payments to control out-of-pocket hospital expenditures in Cambodia', *Health Policy and Planning* **19**(4), 199–208.

Behrman, J. R. & Deolalikar, A. B. (1988), *Health and nutrition*, Elsevier.

Bertrand, M., Duflo, E. & Mullainathan, S. (2002), 'How much should we trust differences-in-differences estimates?', *The Quarterly Journal of Economics* **119**(1), 249–275.

Bolduc, D., Lacroix, G. & Muller, C. (1996), 'The choice of medical providers in rural Benin: A comparison of discrete choice models', *Journal of Health Economics* **15**(4), 477–498.

Bosu, W., Bell, J., Armar-Klemesu, M. & Tornui, J. (2007), 'Effect of delivery care user fee exemption policy on institutional maternal deaths in the Central and Volta Regions of Ghana', *Ghana Medical Journal* **41**(3).

Burnham, G. M., Pariyo, G., Galiwango, E. & Wabwire-Mangen, F. (2004), 'Discontinuation of cost sharing in Uganda', *Bulletin of the World Health Organization* **82**(3), 187–195.

Caliendo, M. & Kopeinig, S. (2008), 'Some practical guidance for the implementation of propensity score matching', *Journal of Economic Surveys* **22**(1), 31–72.

Cameron, A. C. & Trivedi, P. K. (2010), *Microeconometrics using Stata*, Stata Press College Station, TX.

Ching, P. (1995), 'User fees, demand for children's health care and access across income groups: The Philippine case', *Social Science & Medicine* **41**(1), 37–46.

Cogill, B. (2003), 'Anthropometric indicators measurement guide', *Washington, DC: Food and Nutrition Technical Assistance (FANTA) Project* **FHI 360**.

Cohen, J. & Dupas, P. (2007), 'Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment', *Brookings Global Economy and Development Working Paper* (11).

Curtis, S. L. (1995), 'Assessment of the quality of data used for direct estimation of infant and child mortality in DHS-II surveys.', *Occasional Papers* (3).

Deininger, K. & Mpuga, P. (2004), 'Economic and welfare effects of the abolition of health user fees: Evidence from Uganda', *World Bank Policy Research Working Paper* (3276).

Duflo, E. (2002), 'Empirical methods', *Handout of courses MIT* **14**.

Duflo, E. (2003), 'Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in South Africa', *The World Bank Economic Review* **17**(1), 1–25.

Fabricant, S. J., Kamara, C. W. & Mills, A. (1999), 'Why the poor pay more: Household curative expenditures in rural Sierra Leone', *The International Journal of Health Planning and Management* **14**(3), 179–199.

Gertler, P., Locay, L. & Sanderson, W. (1987), 'Are user fees regressive?: The welfare implications of health care financing proposals in Peru', *Journal of Econometrics* **36**(1), 67–88.

Gilson, L. (1997), 'The lessons of user fee experience in Africa', *Health Policy and Planning* **12**(3), 273–285.

Gilson, L. & McIntyre, D. (2005), 'Removing user fees for primary care in africa: The need for careful action', *Bmj* **331**(7519), 762–765.

Graham, W., Brass, W. & Snow, R. W. (1989), 'Estimating maternal mortality: the sisterhood method', *Studies in Family Planning* pp. 125–135.

Hercot, D., Meessen, B., Ridde, V. & Gilson, L. (2011), 'Removing user fees for health services in low-income countries: A multi-country review framework for assessing the process of policy change', *Health Policy and Planning* **26**(suppl 2), ii5–ii15.

Holla, A. & Kremer, M. (2009), 'Pricing and access: Lessons from randomized evaluations in education and health', *Center for Global Development Working Paper* (158).

Hotchkiss, D. R. (1998), 'The tradeoff between price and quality of services in the Philippines', *Social Science & Medicine* **46**(2), 227–242.

International, A. (2011), At a crossroads,: Sierra Leone Free Health Care Policy, Technical report, Amnesty International.

James, C. D., Hanson, K., McPake, B., Balabanova, D., Gwatkin, D., Hopwood, I., Kirunga, C., Knippenberg, R., Meessen, B., Morris, S. S. et al. (2006), 'To retain or remove user fees?', *Applied Health Economics and Health Policy* **5**(3), 137–153.

James, C., Morris, S. S., Keith, R. & Taylor, A. (2005), 'Impact on child mortality of removing user fees: Simulation model', *BmJ* **331**(7519), 747–749.

Khandker, S. R., Koolwal, G. B. & Samad, H. A. (2010), *Handbook on impact evaluation: quantitative methods and practices*, World Bank Publications.

Kremer, M. & Miguel, E. (2004), The illusion of sustainability, Technical report, National Bureau of Economic Research.

Lagarde, M. & Palmer, N. (2008), 'The impact of user fees on health service utilization in low-and middle-income countries: How strong is the evidence?', *Bulletin of the World Health Organization* **86**(11), 839–848C.

Lagarde, M. & Palmer, N. (2011), 'The impact of user fees on access to health services in low-and middle-income countries', *The Cochrane Library* .

Litvack, J. I. & Bodart, C. (1993), 'User fees plus quality equals improved access to health care: Results of a field experiment in Cameroon', *Social Science & Medicine* **37**(3), 369–383.

Malik, K. (2014), Human development report 2014, Technical report, United Nations.

McPake, B. (1993), 'User charges for health services in developing countries: A review of the economic literature', *Social Science & Medicine* **36**(11), 1397–1405.

Meessen, B. (2009), 'Removing user fees in the health sector in low-income countries: A policy guidance note for programme managers', *Health Section Working Paper* .

Meessen, B., Gilson, L. & Tibouti, A. (2011), 'User fee removal in low-income countries: Sharing knowledge to support managed implementation', *Health Policy and Planning* **26**(suppl 2), ii1–ii4.

Meessen, B., Hercot, D., Noirhomme, M., Ridde, V., Tibouti, A., Tashobya, C. K. & Gilson, L. (2011), 'Removing user fees in the health sector: A review of policy processes in six sub-Saharan African countries', *Health Policy and Planning* **26**(suppl 2), ii16–ii29.

Meyer, B. D. (1995), 'Natural and quasi-experiments in economics', *Journal of Business & Economic Statistics* **13**(2), 151–161.

Mwabu, G., Ainsworth, M. & Nyamete, A. (1993), 'Quality of medical care and choice of medical treatment in Kenya: An empirical analysis', *Journal of Human Resources* pp. 838–862.

Mwabu, G. M. & Mwangi, W. M. (1986), 'Health care financing in Kenya: A simulation of welfare effects of user fees', *Social Science & Medicine* **22**(7), 763–767.

Nabyonga, J., Desmet, M., Karamagi, H., Kadama, P. Y., Omaswa, F. & Walker, O. (2005), 'Abolition of cost-sharing is pro-poor: Evidence from Uganda', *Health Policy and Planning* **20**(2), 100–108.

Nolan, B. & Turbat, V. (1995), *Cost recovery in public health services in sub-Saharan Africa*, World Bank Publications.

of Health, M. & Sanitation (2009), National health sector strategic plan (2010-2015), Technical report, Government of Sierra Leone.

of Health, M. & Sanitation (2012), 2010 health sector performance report, Technical report, Government of Sierra Leone.

Pearson, M. (2004), 'Issues paper: The case for abolition of user fees for primary health services', *London: DFID Health Systems Resource Centre* .

Penfold, S., Harrison, E., Bell, J. & Fitzmaurice, A. (2007), 'Evaluation of the delivery fee exemption policy in Ghana: Population estimates of changes in delivery service utilization in two regions', *Ghana Medical Journal* **41**(3), 100.

Preker, A. S., Langenbrunner, J. et al. (2005), *Spending Wisely: Buying Health Services for the Poor*, World Bank Publications.

Ravallion, M. (2007), 'Evaluating anti-poverty programs', *Handbook of Development Economics* **4**, 3787–3846.

Ridde, V. & Diarra, A. (2009), 'A process evaluation of user fees abolition for pregnant women and children under five years in two districts in Niger (West Africa)', *BMC Health Services Research* **9**(1), 89.

Ridde, V. & Morestin, F. (2011), 'A scoping review of the literature on the abolition of user fees in health care services in Africa', *Health Policy and Planning* **26**(1), 1–11.

Ridde, V., Richard, F., Bicaba, A., Queuille, L. & Conombo, G. (2011), 'The national subsidy for deliveries and emergency obstetric care in Burkina Faso', *Health Policy and Planning* **26**(suppl 2), ii30–ii40.

Rosenbaum, P. R. & Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), 41–55.

Rosenzweig, M. R. & Wolpin, K. I. (2000), 'Natural natural experiments in economics', *Journal of Economic Literature* pp. 827–874.

Rutstein, S. O. & Rojas, G. (2006), 'Guide to DHS statistics', *Calverton, Maryland: ORC Macro* .

Schwartz, J. B., Akin, J. S. & Popkin, B. M. (1988), 'Price and income elasticities of demand for modern health care: The case of infant delivery in the Philippines', *The World Bank Economic Review* **2**(1), 49–76.

Shaw, R. P. & Griffin, C. C. (1995), 'Financing health care in Sub-Saharan Africa through user fees and insurance.', *Washington DC World Bank* .

Stanton, C., Abderrahim, N. & Hill, K. (1997), *DHS maternal mortality indicators: An assessment of data quality and implications for data use*, Macro International Calverton.

Stanton, C., Abderrahim, N. & Hill, K. (2000), 'An assessment of DHS maternal mortality indicators', *Studies in Family Planning* **31**(2), 111–123.

Tanaka, S. (2014), 'Does abolishing user fees lead to improved health status? Evidence from post-apartheid South Africa', *American Economic Journal: Economic Policy* **6**(3), 282–312.

Wilkinson, D., Gouws, E., Sach, M. & Karim, S. S. A. (2001), 'Effect of removing user fees on attendance for curative and preventive primary health care services in rural South Africa', *Bulletin of the World Health Organization* **79**(7), 665–671.

Witter, S., Armar-Klemesu, M. & Dieng, T. (2008), 'National fee exemption schemes for deliveries: Comparing the recent experiences of Ghana and Senegal', *Studies in Health Services Organisation and Policy series* **24**, 167–198.

Witter, S., Wurie, H. & Bertone, M. P. (2015), 'The free health care initiative: How has it affected health workers in Sierra Leone?', *Health policy and planning* p. czv006.

Xu, K., Evans, D. B., Kadama, P., Nabyonga, J., Ogwal, P. O., Nabukhonzo, P. & Aguilar, A. M. (2006), 'Understanding the impact of eliminating user fees: utilization and catastrophic health expenditures in Uganda', *Social Science & Medicine* **62**(4), 866–876.

Yates, R. (2009), 'Universal health care and the removal of user fees', *The Lancet* **373**(9680), 2078–2081.